

Published:

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

Gene Editing Mediated by Reverse Transcription of a Modular RNA Template Anchored by a Tag gRNA and Uses Thereof

CROSS REFERENCE TO RELATED APPLICATION

5 This application claims the benefit under 35 U.S.C. §119(e) of the earlier filing date of U.S. Provisional Application No. 63/511,710 filed on July 3, 2023. The content of the application is incorporated herein by reference in its entirety.

GOVERNMENT INTERESTS

10 This invention was made with government support under MD200088 awarded by the Department of Defense. The government has certain rights in the invention.

REFERENCE TO AN ELECTRONIC SEQUENCE LISTING

15 The contents of the electronic sequence listing (096738.00780SeqList.xml; Size: 162,826 bytes; and Date of Creation: June 28, 2024) is herein incorporated by reference in its entirety.

FIELD OF THE INVENTION

This disclosure relates to gene editing, related systems, and uses thereof.

BACKGROUND

20 Targeted gene editing has been used for genetic manipulation of eukaryotic cells, embryos, and animals. Sequence-specific nucleases, including Talen, Zinc finger nucleases, and RNA- guided nucleases such as CRISPR/Cas9, have provided tools for precision genome editing. Upon binding to the target sequence, the native nuclease generates a DNA double-strand break (DSB), eliciting cellular DNA repair pathways including non-homologous end joining (NHEJ) and homology-directed repair (HDR). As a result, a desired sequence change, 25 or gene editing, could be achieved.

In the editing process, the on-target and off-target DNA DSB intermediates can lead to chromosomal translocation and other mutagenic events with potential oncogenic liability. To avoid the requirement of DNA DSBs, base editing platforms were developed. The CRISPR base editors use a nuclease-null or nickase version of CRISPR proteins. Proficient in DNA 30 sequence recognition, the mutant CRISPR proteins do not cause DSBs. Instead, the mutant CRISPR protein-gRNA complex recruits a cytidine deaminase or adenine deaminase, which in

turn converts a C to U or A to G at the target, respectively, leading to sequence-specific point mutations. Avoiding DSBs, base editing reduces the oncogenic liability and is broadly used for precision genome editing for both basic research and therapeutic development.

As nucleotide deamination is the underlying mechanism for base change, base editors edit transition point mutations but not transversion mutations. In addition, base editors require the target editing nucleotide within the R-loop adjacent to the PAM motif. The requirement excludes the accessibility of target sites that do not have an adjacent PAM motif. Moreover, within the editing activity window, the deamination is generally promiscuous, which causes bystander base editing within the window. While bystander base editing may be innocuous in some therapeutic developments, such as correcting a loss-of-function mutation, precision editing free of bystander editing is generally preferred under many other circumstances.

Prime editing is another precision gene editing platform that does not require double strand DNA break (Anzalone, A.V., *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019)). The Prime Editor complex comprises a nickase version of CRISPR protein fused with a reverse transcriptase (RT) and a modified gRNA called pegRNA containing an RNA template and a primer binding sequence, commonly at the 3' end of gRNA. After binding to the target DNA sequence, the Prime Editor creates a DNA nick. The nicked DNA strand serves as a primer, which binds to the primer binding sequence within the pegRNA and synthesizes a new DNA sequence using the RNA template within the pegRNA. The sequence information of the template RNA is then copied from RNA to DNA by the reverse transcriptase. In turn, the DNA sequence is further incorporated into the target site. By copying a desired RNA template at the 3' end of pegRNA, Prime Editing is capable of achieving bystander-free precision genome editing of base pair change. The base change can be both transition and transversion. Moreover, by designing insertion and deletion in the pegRNA template, Prime Editing can also create insertion and deletion at the target location.

At the center of Prime Editing system is the modified gRNA named pegRNA, which contains both gRNA scaffold, a primer binding sequence, and an editing template in the same RNA molecule. This configuration, however, limits the size of the editing template, as a long RNA template sequence within the same RNA molecule of the gRNA would create secondary structures that may interfere with the secondary structure of the gRNA CRISPR protein binding scaffold, for example. The longest template tried in Anzalone, A.V., *et al.* *Nature* **576**, 149–157 (2019) within pegRNA molecules is 34 nt in length. The template size restriction makes

some target sites in the genome inaccessible by Prime Editors due to the lack of proper PAM motif nearby. Even if a PAM motif is within the limit of the RT (reverse transcriptase) template size limit, sometimes the template size limit may restrict the optionality of RT template selection, as the sequence adjacent to PAM may have suboptimal GC content, making some target sites in the genome impractical for prime editing due to unfavorable GC content for priming and extension.

One variant configuration of Prime Editing separates pegRNA into two molecules, namely an unmodified gRNA and an RT template with an exogenous appendicular RNA aptamer structures, *e.g.*, MS2 (WO2020/191248A1). In this configuration, the CRISPR protein needs to contain an additional fusion partner for interacting with the RNA aptamer within the RT template, *e.g.*, an MCP protein is fused to Cas9 protein for recruiting the MS2 RNA aptamer-containing RT template. Within the system, the gRNA is responsible for complexing with CRISPR protein for target site recognition. The RNA aptamer such as MS2 is responsible for being recruited to the CRISPR complex by the additional fusion moiety (*e.g.*, MCP). The Prime Editors with this configuration exhibit editing efficiency that is a fewer fold lower than their pegRNA Prime Editor counterparts (Figure 73 of WO2020/191248A1). Moreover, the ratios of undesired insertion/deletion vs. correct editing are higher than pegRNA Prime Editor. WO2020/191248A1 also showed that the inclusion of appendicular RNA aptamer in the RT template, in particular at the 5' end, would potentially lead to copying the unwanted exogenous genetic information (RNA aptamer) into the target site.

There is a need for more systems that are versatile and efficacious for editing a target nucleic acid molecule.

SUMMARY

This disclosure addresses the need mentioned above in a number of aspects.

In one aspect, the disclosure provides a gene editing complex for editing a target site in a target DNA molecule. The gene editing complex comprises (A) an RNA-guided nickase; (B) a reverse transcriptase; (C) an RNA template molecule, and (D) a matching gRNA (magRNA) molecule. The RNA template molecule comprises (1) a template segment comprising a template sequence complementary to a desired DNA sequence or a complement thereof to be introduced into the target site, and (2) a priming segment complementary to the 3' end of the nicking strand of the target DNA molecule. The magRNA molecule comprises (1) a guide or spacer sequence complementary to a sequence on a target strand of the target DNA molecule,

(2) an RNA scaffold capable of binding to the RNA-guided nickase, and (3) an anchoring tag sequence complementary to a segment in the RNA template.

The anchoring tag and the matching tag are used interchangeably in the application. Both are referred to the tag sequence added to the gRNA that complementarily matches to a region in the modular RNA template for recruiting and anchoring the RNA template.

The reverse transcriptase can be incorporated into the complex in any suitable means. In one embodiment, the reverse transcriptase is linked (covalently or non-covalently) to or fused to the RNA-guided nickase. In another embodiment, the magRNA further comprises a protein-binding motif (*e.g.*, MS2 or PP7) capable of binding to an RNA-interacting protein, and the reverse transcriptase is linked (covalently or non-covalently) to or fused to the RNA-interacting protein (*e.g.*, MCP or PCP).

In one embodiment, the anchoring tag in the magRNA is not a polyN, wherein N is a repeating nucleotide A, C, U, or G.

In one embodiment, the RNA template does not comprise additional sequence other than the said priming segment and template segment.

In one embodiment, the RNA-guided nickase is a nickase variant of a Cas protein or a nickase variant of transposon-encoded RNA-guided endonuclease protein. In one embodiment, the Cas protein is selected from Cas3, Cas4, Cas5, Cas5e (or CasD), Cas6, Cas6e, Cas6f, Cas7, Cas8a1, Cas8a2, Cas8b, Cas8c, Cas9, Cas10, Cas10d, Cpf1 (Cas12a), C2c1 (Cas12b), C2c3 (Cas12c), CasY (Cas12d), CasX (Cas12e), Cas14 (Cas12f), CasPhi (Cas12j), Cas13a, Cas13b, Cas13c, Cas13d, Cas13x, CasF, CasG, CasH, Csy1, Csy2, Csy3, Cse1 (or CasA), Cse2 (or CasB), Cse3 (or CasE), Cse4 (or CasC), Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csz1, Csx15, Csf1, Csf2, Csf3, Csf4, Cu1966, and the orthologues thereof. In one embodiment, the transposon-encoded endonuclease proteins are the IscB, IsrB, or TnpB endonuclease proteins and the orthologues thereof.

In one embodiment, the RNA-guided nickase is a nickase variant of an ortholog of Cas protein or a nickase variant of ortholog of transposon-encoded RNA-guided endonuclease protein listed above. In one embodiment, the Cas protein is Cas9. In one embodiment, the nickase variant of Cas9 ortholog protein is *Streptococcus pyogenes* (*sp*) nCas9(H840A) or *Staphylococcus aureus* (*sa*) nCas9 (N580A). In one embodiment, the nickase variant of Cas9 ortholog protein is *Staphylococcus aureus* nCas9 (N580A E782K/N968K/R1015H).

In one embodiment, the reverse transcriptase is a naturally-occurring reverse transcriptase from a retrovirus or a retrotransposon, or a variant thereof. In one embodiment, the reverse transcriptase is selected from the group consisting of Moloney Murine Leukemia Virus (M-MLV), Human Immunodeficiency Virus (HIV) reverse transcriptase, Avian Sarcoma-Leukosis Virus (ASLV) reverse transcriptase, Rous Sarcoma Virus (RSV) reverse transcriptase, Avian Myeloblastosis Virus (AMV) reverse transcriptase, Avian Erythroblastosis Virus (AEV) Helper Virus MCAV reverse transcriptase, Avian Myelocytomatosis Virus MC29 Helper Virus MCAV reverse transcriptase, Avian Reticuloendotheliosis Virus (REV-T) Helper Virus REV-A reverse transcriptase, Avian Sarcoma Virus UR2 Helper Virus UR2AV reverse transcriptase, Avian Sarcoma Virus Y73 Helper Virus YAV reverse transcriptase, Rous Associated Virus (RAV) reverse transcriptase, and Myeloblastosis Associated Virus (MAV) reverse transcriptase, Line 1 ORF2, R2Bm, and R2O1. In one embodiment, the reverse transcriptase is an MMLV-RT or a variant thereof.

In one embodiment, the retroviral and retrotransposon derived reverse transcriptases have intrinsic DNA-dependent DNA polymerase activity, *e.g.* Line 1 ORF 2 and HIV RT. In one embodiment, the retrotransposon derived reverse transcriptase is R2Bm.

In one embodiment, the anchoring tag sequence can be at the 3' or 5' end of the magRNA. In one embodiment, the anchoring tag sequence is covalently linked to the magRNA via a polynucleotide linker. In one embodiment, the anchoring tag sequence in the magRNA is 6-24 nucleotide in length.

In one embodiment, the anchoring tag sequence of magRNA is complementary to the 5' end of the RNA template molecule. In one embodiment, the anchoring tag sequence of magRNA is complementary to an internal region of the RNA template.

In a second aspect, the disclosure features a system for editing a target site in a target DNA molecule. The system comprises (I) a first gene editing complex of any one of those described above, and (II) a second gene editing complex.

In one embodiment, the second gene editing complex comprises (A) a second RNA-guided nickase, and (B) a gRNA molecule comprising a second guide or spacer sequence, with or without an RNA aptamer (*e.g.*, MS2).

In one embodiment, the second gene editing complex comprises (A) a second RNA-guided nickase; and (B) a second magRNA molecule comprising (1) a second guide or spacer sequence, (2) a second RNA scaffold capable of binding to the second RNA-guided nickase, and (3) a second anchoring tag sequence.

In one embodiment, the second gene editing complex comprises (C) a second reverse transcriptase.

In one embodiment, the two guide sequences in the two gene editing complex (*i.e.*, the guide sequence of the magRNA in the first gene editing complex and the second guide
5 sequence of the gRNA or the second magRNA in the second gene editing complex) are complementary to two target sequences of the target DNA molecule, respectively.

In one embodiment, the second anchoring tag sequence of the second magRNA is complementary to a second segment within the RNA template.

In one embodiment, the 5' end of the RNA template comprises a sequence identical to
10 the 3' end of the strand of the target DNA molecule that is nicked by the second RNA-guided nickase.

In one embodiment, the second gene editing complex further comprises (C) a second reverse transcriptase, or (D) a second RNA template, or both.

The second reverse transcriptase can be incorporated into the second gene editing
15 complex or the system in any suitable means. In one embodiment, the second reverse transcriptase can be linked (covalently or non-covalently) to or fused to the second RNA-guided nickase. In one embodiment, the second magRNA molecule or the gRNA molecule comprises a second protein-binding motif capable of binding to a second RNA-interacting protein, and the second reverse transcriptase is linked (covalently or non-covalently) to or fused
20 to the second RNA-interacting protein.

In one embodiment, the second gRNA or magRNA in the second gene editing complex contains an RNA aptamer (*e.g.* MS2 or PP7) which recruits a non-reverse transcriptase effector fused with the RNA aptamer binding protein (*e.g.*, MCP or PCP). In one embodiment, the non-reverse transcriptase effector is the 50-kDa RNase Inhibitor 1 protein (RNH1) or its orthologue.
25 In one embodiment, the non-reverse transcriptase effector is a dominant negative protein of MMR pathway protein including MLH1 protein, or a 5' DNA nuclease Fen1 protein.

In one embodiment, the non-reverse transcriptase effector is a pioneer transcription factor. In one embodiment, the pioneer transcription factor is p65.

In one embodiment, the non-reverse transcriptase effector is a chromatin modulating
30 protein or protein complex. In one embodiment, the chromatin modulating protein has histone acetylation activity.

Also provided in this disclosure is a method of modifying a target DNA molecule in a cell. The method comprises contacting the target DNA molecule with the above-described gene

editing complex or the above-described system. In one embodiment, said modifying leads to a point mutation to the target DNA molecule, an insertion to the target DNA molecule, a deletion to the target DNA molecule, or a combination thereof. In one embodiment, the cell is selected from the group consisting of: an archaeal cell, a bacterial cell, a eukaryotic cell, a eukaryotic single-cell organism, a somatic cell, a germ cell, a stem cell, a plant cell, an algal cell, an animal cell, in invertebrate cell, a vertebrate cell, a fish cell, a frog cell, a bird cell, a mammalian cell, a pig cell, a cow cell, a goat cell, a sheep cell, a rodent cell, a rat cell, a mouse cell, a non-human primate cell, and a human cell. In one embodiment, the cell is in, isolated from, or derived from a human or non-human subject.

In a further aspect, this disclosure provides a genetically engineered cell obtained according to the method described above or a progeny thereof. In one embodiment, the cell is selected from the group consisting of an archaeal cell, a bacterial cell, a eukaryotic cell, a eukaryotic single-cell organism, a somatic cell, a germ cell, a stem cell, a plant cell, an algal cell, an animal cell, in invertebrate cell, a vertebrate cell, a fish cell, a frog cell, a bird cell, a mammalian cell, a pig cell, a cow cell, a goat cell, a sheep cell, a rodent cell, a rat cell, a mouse cell, a non-human primate cell, and a human cell. In one embodiment, the cell is selected from the group consisting of pluripotent stem cells (PSCs), adult stem cells (ASCs), hematopoietic stem cells (HSCs) fibroblasts, chondrocytes, keratinocytes, hepatocytes, pancreatic islet cells, and immune cells, including T cells, dendritic cells (DCs), natural killer (NK) cells, and macrophages isolated or derived from a human or non-human subject.

In another aspect, the disclosure features a magRNA as described above. The magRNA comprises (1) a guide or spacer sequence complementary to a target sequence on a target strand of the target DNA molecule; (2) an RNA scaffold capable of binding to the RNA-guided nickase; and (3) an anchoring tag sequence complementary to a segment in the RNA template.

The magRNA may further comprise a protein-binding motif capable of binding to an RNA-interacting protein, and the reverse transcriptase is linked to the RNA-interacting protein. In one embodiment, the anchoring tag in magRNA is not a polyN, wherein N is a repeating nucleotide A, C, U, or G. In one embodiment, the anchoring tag sequence is at the 3' or 5' end of the magRNA. In one embodiment, the anchoring tag sequence is covalently linked to the magRNA via a polynucleotide linker. In one embodiment, the anchoring tag sequence in the magRNA is about 6-24 nucleotide in length.

In another aspect, the disclosure features an RNA complex comprising the magRNA described above and the magRNA- anchored RNA template described above.

The anchored RNA template does not contain any extra sequence other than the desired sequence designed to be present in the edited genome.

In another aspect, the disclosure provides a nucleic acid encoding one or two of: (i) the magRNA molecule described above, and (ii) the RNA complex described above.

5 In another aspect, the disclosure features a vector comprising the nucleic acid.

In another aspect, the disclosure features a kit comprising (i) a packaging material, and (ii) one, two, or more of: the above-described gene editing complex, the above-described system, the above-described cell, the above-described magRNA molecule, the above-described RNA complex, the above-described nucleic acid, and the above-described vector.

10 In another aspect, the disclosure provides a pharmaceutical composition comprising (i) a pharmaceutically acceptable carrier, and (ii) one, two, or more of: the above-described gene editing complex, the above-described system, the above-described cell, the above-described magRNA molecule, the above-described RNA complex, the above-described nucleic acid, and the above-described vector.

15 The details of one or more embodiments of the disclosure are set forth in the description below. Other features, objectives, and advantages of the disclosure will be apparent from the description and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGs. 1A, 1B, and 1C show a single-module Match Editing system using
20 CRISPR/gRNA for RNA-guided sequence-recognition. The system comprises (FIG. 1A) a modular RNA template that does not contain any exogenous RNA recruiting elements such as RNA aptamers (*e.g.*, MS2 or PP7); (FIG. 1B) a magRNA containing a 5' spacer sequence for target DNA recognition, an RNA scaffold for CRISPR/Cas complexing, and a 3' match RNA sequence that is complementary to a region in the modular template; and (FIG. 1C) a nickase
25 CRISPR protein, nCRISPR/nCas9(H840A), fused with a reverse transcriptase (MMLV-RT) through a polypeptide linker.

FIG. 2 shows how a Match Editing system works. nCRISPR/nCas9-RT forms a complex with magRNA. The 5' spacer sequence of magRNA recognizes and complements with a target sequence in the target DNA; in the meantime, the 3' tag sequence of magRNA
30 complements with a region in the RNA template, anchoring the RNA template to the CRISPR-RT/magRNA complex. The nickase activity of the nCRISPR/nCas9 leads to a single-strand DNA break (nicking) upstream of the PAM motif. The nicking DNA then binds to the 3' end of RNA template through sequence complementation, further strengthening the complex. The

free 3' strand of the nicking DNA then serves as primer for the reverse transcriptase, using the complexing RNA as the nucleotide extension template.

FIG. 3 shows that a cellular repair mechanism leads to incorporation of a newly synthesized DNA strand into a target DNA locus. FIG. 3A shows the relative locations of the RNA-guided nicking site (showing as a star in the upper DNA strand, indicated with “nicking”) and a desired mutation site (showing as complementary arrows in the DNA strands, indicated with “mutations”). FIG. 3B-D show the synthesis of new DNA strand extending the 3' nicking strand, followed by removing the 5' flapping original DNA fragment, resulting in a new upper strand with desired sequence change and a nick. FIG. 3E shows the resolving of the mismatch and filling the nicking gap by cellular repair mechanisms, resulting in either a DNA molecule with the desired edits (by removing the mismatch at the lower strand) or a DNA of the original sequence (by removing the mismatch at the upper strand).

FIG. 4 shows an example of magRNA-RNA template interaction, where the 3'-matching tag sequence of a magRNA is complementary specifically to the 5'-end sequence of an RNA template.

FIG. 5 shows a magRNA with a 5' match tag that is complementary to a region of an RNA template and is located at the 5' end of a gRNA. A linker polynucleotide sequence is located between the 5' match tag sequence and the spacer sequence. The mechanism of editing is similar as described in FIG. 2, except for the interaction position between the tag sequence of the magRNA and the template RNA.

FIGs. 6A, 6B, 6C, and 6D show a variant Match Editing system where a reverse transcriptase (RT) is provided *in split* rather than directly fused to an RNA-guided sequence specific nickase.

FIG. 6A shows a modular RNA template.

FIG. 6B shows an engineered magRNA. The magRNA contains a 5' spacer sequence for target sequence recognition, a 3' tag sequence for recruiting an RNA template (FIG. 6A), and an RNA aptamer (*e.g.*, MS2) at the gRNA stem loop region (SL) for recruiting the reverse transcriptase individually.

FIG. 6C shows an RNA guided nickase without a directly linked reverse transcriptase.

FIG. 6D shows a protein comprising a reverse transcriptase fusing with a cognate protein (*e.g.*, MCP) binding to an aptamer within a gRNA scaffold (*e.g.*, MS2) through a polypeptide linker.

FIGs. 7A, 7B, 7C, and 7D show how a reverse transcriptase in an *in split* system works.

FIG. 7A shows a modular RNA template.

FIG. 7B shows an engineered magRNA.

FIG. 7C shows an RNA guided nickase.

FIG. 7D shows a protein comprising a reverse transcriptase fusing with a cognate
5 protein (*e.g.*, MCP) binding to an aptamer within a gRNA scaffold (*e.g.*, MS2) through a polypeptide linker. The process is similar to the description in FIG. 2, except for the RT is recruited in split.

FIG. 8 shows a variant configuration of RT *in split* system, where a 3' magRNA extension sequence is complementary to a sequence specifically at 5' end of an RT template.

10 FIGs. 9A and 9B show examples of Dual ME systems.

FIG. 9A shows a Dual ME system with fusing RT and with magRNA-gRNA pairing. In this configuration, the first ME contains typical ME components including a RT fused with nCRISPR/nCas9 and a magRNA. The second module is for creating a second nick for increasing editing efficiency. The gRNA in the second module does not have a 3' tag sequence
15 for RNA template interaction.

FIG. 9B shows a Dual ME system with trans RT and magRNA with MS2-gRNA 0xMS2. In this configuration, the first ME contains a magRNA with an aptamer for recruiting the reverse transcriptase in split. The second module is for creating a second nick for increasing editing efficiency. The gRNA of the second ME module does not contain a 3' matching tag for
20 RNA template recruiting.

FIG. 10 shows that a second nicking at the nearby opposite strand mediated by a dual ME system greatly enhances target editing efficiency. FIGs. 10A to 10C is the same as FIG. 3A-C. FIG. 10D shows a second nicking site is created at the lower strand, which favors the removal of the DNA flap containing a free 5' nicking end over the DNA flap containing a free
25 3' nicking end. As a result, the reverting to the original DNA sequence is disfavored (FIG. 10E, right, with stop signs), and the products are predominantly DNA molecules with the desired edits (FIG. 10E, left).

FIGs. 11A and 11B show a Single Template-Coupled Dual ME system for gene editing with a large DNA fragment. FIG. 11A shows a Coupled Dual ME system with a fused RT.
30 FIG. 11B shows a Coupled Dual ME system with a RT *in split*. A large reverse transcriptase RNA template is supplied in this configuration, to allow editing of the desired sequence at location farther away from the RNA-guided nicking site generated by the first ME module. In this configuration, the second ME module contains a magRNA with a 3' tag sequence

interacting with the RNA template at a different segment. The difference between FIG. 11A and 11B is that the reverse transcriptase is covalently linked to nCRISPR/nCas9 in 11A, while the RT is supplied individually in 11B.

FIGs. 12A and 12B show exemplary versions of Coupled Dual ME systems with a second priming mechanism.

FIG. 12A shows a RT fusion version, where the 5' end of an RNA template is designed as identical to the 3' of the second nicking strand. As a result, the second nicking strand serves as the primer for second DNA strand synthesis using the newly synthesized DNA strand as a template.

FIG. 12B shows a RT-*in split* version.

FIG. 13 shows the scheme of large fragment insertion involving second strand DNA-dependent DNA synthesis by the single-template coupled Dual ME system. FIG. 13A shows the first RNA guided nicking by the first ME module. FIG. 13B shows the first DNA strand synthesis by the first ME module using modular RNA template. FIGs. 13C-D show the introduction of the second nicking by the second ME module. FIG. 13E shows the priming by the free 3' end of the second nicking strand with the newly synthesized DNA strand. FIG. 13F shows the strand extension and synthesis of the second DNA strand via the DNA-dependent DNA polymerase activity of the reverse transcriptase or via the cellular DNA-dependent DNA polymerase. FIGs. 13G-H show the completion of the insertion of the newly synthesized DNA fragment by removing the two DNA flaps containing free 5' ends and subsequent ligation via cellular repair mechanism.

FIG. 14 shows a Dual Template-Dual Module system for insertion and/or deletion. In this configuration, both ME modules contain a magRNA and RNA template. The reverse transcriptase can fuse to the nCRISPR/nCas9 or supply in split.

FIGs. 15A, 15B, 15C, 15D, 15E, and 15F show how a dual-template dual-module ME leads to *de novo* gene composition (insertion) or large fragment gene deletion.

FIGs. 15A and 15B show module 1 ME and module 2 ME mediate nicking at their corresponding target sites and lead to first strand DNA synthesis by reverse transcription.

FIG. 15C shows annealing of the 3' ends of the two newly synthesized first DNA strands through the complementary sequences.

FIG. 15D shows second strand DNA synthesis by DNA-dependent DNA polymerase activity of the reverse transcriptase or endogenous cellular DNA polymerase activity.

FIGs. 15E and 15F show removal of 5' flaps and ligation. The mechanism can lead to both insertion and deletion (*de novo* gene composition) depending on the content and length of the two RNA templates and the sequences removed.

FIGs. 16A-D show ME systems with second effectors.

5 FIG. 16A shows modular RNA template.

FIG. 16B shows a magRNA with a 5' spacer sequence for target site recognition, a 3' tag sequence for reverse transcription RNA template recruitment, an aptamer sequence (*e.g.*, MS2) at the stem loop location for recruitment of a second effector.

10 FIG. 16C shows the RNA guided nickase (*e.g.*, nCRISPR/nCas9) fused with a reverse transcriptase (*e.g.*, MMLV-RT) through a linker peptide.

FIG. 16D shows a second effector protein (*e.g.*, RNase inhibitor or FEN1) fused to the aptamer binding protein (*e.g.*, MCP) through a peptide linker.

FIGs. 17A, 17B, 17C, and 17D show Match Editing System components for correcting the point mutation A200G in a nfEGFP gene.

15 FIG. 17A shows expression plasmids of components of a Match Editing system, and a gRNA control.

FIG. 17B shows the nfEGFP gene to be edited.

FIG. 17C shows a sequence of a target nicking site (SEQ ID NOs: 1 and 2), showing PAM (underlined) and nicking position (arrow); the mutation G200 is boxed.

20 FIG. 17D shows a Match Editing template having a primer binding sequence (P14, SEQ ID NO:3) and an RT extension with a length of 29 nt or 57 nt.

FIGs. 18A, 18B, and 18C show efficacy of the Match Editing system on correcting a point mutation shown in FIG. 17.

25 FIG. 18A shows results of the Match Editing system on correcting a point mutation by a functional assay.

FIGs. 18B and 18C show Sanger sequencing results (SEQ ID NO: 4) from Lane 2 and Lane 4 of FIG. 18A, respectively. A shorter template with 29nt RT extension (Ex29) was used for Lane 3 and 4, with a gRNA or magRNA, as indicated. A longer template was used for Lane 5 and 6, with a gRNA or magRNA, as indicated. The tag sequences of the mag RNAs are 12-
30 nt in length, matching the 5' end sequences of the corresponding template. PEG RNA mediated Prime Editing (Lane 7) was used for comparison.

FIGs. 19A and 19B show comparison of editing efficacies of magRNAs matching an internal sequence (12M29) vs 5'-sequence (12M57) of a long template Ex57.

FIG. 19A shows results of fluorescent microscopy.

FIG. 19B shows percentages of cells expressing fluorescent EGFP as determined by flow cytometry. The Panels 1 to 7 of FIG. 19A are the same corresponding panels in FIG. 19B. For each panel, the treatments are indicated in FIG. 19B.

FIGs. 20A, 20B, and 20C show efficacy of ME in correcting deletion mutations. The Panels 1 to 5 in FIG. 20A are the same corresponding Panels in FIG. 20B.

FIG. 20A shows fluorescent microscopy results.

FIG. 20B shows percentages of cells expressing fluorescent EGFP as determined by flow cytometry. Panel 1, untreated cells; Panel 2, cells electroporated with EGFP deletion mutants (an equal molar mixture of $\Delta 4A$, $\Delta 20C$, and $\Delta 35C$). Panels 3, 4, and 5, cells expressing ME components with indicated EGFP deletion, template, and magRNA (12M29).

FIG. 20C shows DNA sequencing results (SEQ ID NO: 5) from Panel 3 cells.

FIG. 21 shows comparison of editing efficacies of magRNAs matching an internal (12M29) vs. 5' end sequence (12M57) of a long template Ex57 for insertion editing. Panel 1, untreated cells; Panel 2, cells electroporated with an equal mixture of EGFP deletion plasmid $\Delta 35C$ and $\Delta 50C$; Panels 3-6, cells expressing ME components with indicated EGFP plasmids, templates, and magRNA.

FIG. 22 shows an insertion editing rate of ME at position 100 nucleotide from a target nicking site (SEQ ID NO: 6), using a template with 111 nt RT extension length, a magRNA matching the internal sequence (12M29).

FIG. 23A compares Match Editing efficiency in correcting a deletion mutation in EGFP gene with magRNAs with varying matching lengths (9nt to 24 nt).

FIG. 23B compares Match Editing efficiency in correcting a deletion mutation in EGFP gene with magRNAs with the same length (12 nt) but varying matching positions (3' tag position complementary with positions 23nt to 41nt upstream (5') of template relative to RT initiation site).

FIGs. 24A, 24B, and 24C show a Trans RT Match Editing system, its components, and its efficiency in correcting a deletion in EGFP.

FIG 24A shows the Trans RT ME system and components thereof where the reverse transcriptase is not fused with nCas9. The magRNA contains an MS2 aptamer at the stem loop position. An RT is fused to MCP through a linker peptide.

FIG. 24B shows a gene to be edited, an EGFP gene containing a C deletion at the position 50 nucleotides upstream of a 200-L nicking site.

FIG. 24C shows efficacy of the Trans RT Match Editing system in correcting the deletion in comparison with the direct fusion ME system.

FIGs. 25A, 25B, and 25C show a Dual ME system with magRNA-gRNA pairing and its efficacy on editing.

5 FIG. 25A shows components of the Dual ME system.

FIG. 25B shows two second-module sites tested (SEQ ID NOs: 7 and 8). The second module ME contains a gRNA instead of a magRNA.

FIG. 25C shows editing efficacies of a single module ME and the Dual MEs.

10 FIGs. 26A, 26B, and 26C show that Match Editing effectively changed a G to A at an endogenous site, HEK4 site.

FIG. 26A shows results from non-treated cells (SEQ ID NO: 9).

FIG. 26B shows results from cells electroporated with an ME system with the target site of HEK4 containing a gRNA without a matching tag (SEQ ID NO: 9).

15 FIG. 26C shows results from cells electroporated with an ME system with the target site of HEK4 containing a magRNA with a matching tag to template (SEQ ID NO: 9).

FIG. 27 shows that Match Editing was effective in deletion editing. The mutant EGFP gene contains a 4-nt insertion starting at position 180. The ME template with 29-nt RT extension length was used for ME editing, together with the 12M29 magRNA containing a 12-nt matching tag to the 5' end of template. A Prime Editing experiment with pegRNA was
20 performed as a control.

FIGs. 28A and 28B compared the efficiency of deletion editing by a single-module ME, a dual-module ME, a single module Primer Editor (PE2), and a dual module PE (PE3), in deleting the splicing donor site (SDS) of mouse dystrophin gene exon 23 in a reporter construct, leading to exon 23 skipping.

25 FIG. 28A shows a schematic of a GFP based splicing reporters.

The upper expression construct contains an EGFP gene split into a 5' half and a 3' half, interrupted by an artificial intron containing splicing donor sequence (SDS) and splicing acceptor sequence (SAS). Transcription and splicing of the splicing reporter gene will lead to the production of a functional EGFP.

30 The lower expression vector is constructed by inserting the mouse dystrophin gene exon 23 with intron 22 and intron 23 splicing regulatory sequences. Transcription and splicing of the reporter gene will lead to the non-fluorescent production of EGFP protein with an insertion

peptide encoded by exon 23. Deleting Dmd Exon 23 SAS or SDS will lead to Exon 23 skipping, resulting the production of a fluorescent EGFP.

A single-module ME and a dual-module ME with magRNA-gRNA pairing were designed for deleting a 29 nt sequence at the Dmd exon 23-intron 23 juncture, eliminating the SDS sequence, as indicated. The Primer Editor PE2 and PE3 with the same RT extension length and primer binding sites were designed for comparison. The nicking site for the first ME or PE2 (1st nicking) and the second nicking site for dual ME or PE3 are indicated by arrows.

FIG. 28B shows results from cells that were either untreated, or electroporated with mDmd exon 23 splicing reporter alone, or the reporter with a single module ME (magLow), or a dual-module ME (magLow+sg), or PE2, or PE3. A splicing reporter served as a positive control.

FIG. 29 compares the editing efficiency of a single-module ME, a dual-module ME with magRNA-gRNA pairing, and a dual-module ME with magRNA-magRNA pairing. The ME systems are similar to those illustrated in FIG. 25, except for the last panel experiment, wherein the second gRNA also contains a second matching tag at 3' end. The results show cells that were either untreated, or electroporated with an mutant EGFP (deletion of one nucleotide at position 50 relative to nicking site) expression vector (EGFP Δ50), or with the mutant EGFP expression vector plus a one-module ME (magLow), or a dual-module ME with magRNA-gRNA pairing (magLow+sgUp119), or a dual-module ME with magRNA-magRNA pairing (magLow +magUp119), all with an RNA template of an extension length of 111 nt. In the magRNA-magRNA pairing, the match tag sequences in the two magRNAs are complementary to two separate sequences in the RNA RT template.

FIGs. 30A, 30B, and 30C show editing efficiency at the endogenous HEK3 site by dual ME with varying ratios of RNA template, magRNA, and second gRNA, in HEK293T cells.

FIG. 30A illustrates the target site, HEK3, sequence (5' TGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAGCCCTGCTTCCTCCAGAGGGC GTCGCAGGACAGCTTTTCCTAGACAGGGGCTAGTATGTGCAGC, SEQ ID NO: 10) and its complement sequence (5' GCTGCACATACTAGCCCCCTGTCTAGGAAAAGCTGTCCT GCGACGCCCTCTGGAGGAAGCAGGGCTTCCTTTTCCTCTGCCATCACGTGCTCAGTCTGGGCC CCA, SEQ ID NO: 11) annotated with the dual magRNA-gRNA ME system components, including magRNA (12M42) guide, 2nd nicking gRNA guide, and desired point mutations, 5G>T and 12 G>C, encoded in the RNA template T43. +63 indicates the distance between the

two nicks. For illustration purposes, only the 5' end and 3' end sequences of the target site are shown.

FIG. 30B shows representative Sanger sequencing chromatographs of the untreated (UT) and edited (2 mut) HEK3 genomic DNA (SEQ ID NO: 12).

5 FIG. 30C compares the editing efficiency (5G>T) at the endogenous genomic target, HEK3, under conditions with varying ratio between RNA template (T43), magRNA, and second gRNA.

FIGs. 31A, 31B, and 31C show insertion editing efficiency at the endogenous HEK3 site in HEK 293T cells by a dual ME system.

10 FIG. 31A shows the HEK3 sequence and its complement sequence (SEQ ID NOs: 10 and 11) with annotations of magRNA (12M42) guide, second nicking guide, nicking sites, PAMs, as well as the insertion position (+5). The RNA templates (T46, T50, and T79) were designed to insert a 3-nt (CTT), 7-nt AP1 site (ACTCAGT), or a 36-nt TCR variable region, into the endogenous HEK3 genomic site. For illustration purposes, only the 5' end and 3' end
15 sequences of the target site are shown.

FIG. 31B shows representative Sanger sequencing chromatographs of the untreated (UT) and edited genomic DNA (SEQ ID NO: 13).

FIG. 31C shows the efficiency of 3nt, 7nt, and 36nt insertions. Even though the nCas9-RT fusion was used in this experiment, the second gRNA construct also contained an MS2 at
20 the stem-loop region.

FIGs. 32A, 32B, and 32C show the editing efficiency of a trans dual ME system on correcting a point mutation in the episomal nfEGFP reporter gene in HEK293 cells, when various polymerases, including MMLV-RT, R2Bm (a retrotransposon RT), and a variant of Helraiser (a polymerase for DNA transposon Heliton) were recruited. An RNA and a DNA
25 template were tested when Helraiser was used. The trans ME system configuration was previously shown in FIG. 24A. FIG. 32A shows the components of the Trans ME system. The gene to be edited is the nfEGFP target site with mutation of A200G (FIG. 32B). FIG. 32C shows the editing efficacy of the trans dual ME under various conditions, quantified by the percentage of cells with edited, fluorescent EGFP.

30 FIGs. 33A and 33B show the effect of recruiting a non-RT effector p65 and the presence of an MS2 in the second gRNA on editing efficiency in HEK293T cells. FIG. 33A shows the dual magRNA-gRNA ME components. The endogenous target site, HEK3, was previously

shown in FIG. 30A. FIG. 33B compares the editing efficacy (5G>T, Sanger sequencing) of the dual ME systems with or without p65, with the second gRNA with or without an MS2.

FIGs. 34A, 34B, 34C, and 34D show results of assays that compare the editing efficacy between a split dual ME system and its counterpart split prime editing system (PE) on the HEK3 site in HEK293T cells. The target HEK3 sequence was previously shown in FIG. 30A.

FIGs. 34A-C shows the components of ME and PE. FIG. 34A-B show ME- and PE-specific components, respectively. The pegRNA contains the same RNA template sequence of ME, covalently linked to the 3' of gRNA. FIG. 34C shows the common components of the trans ME and PE systems. FIG. 34D shows the editing efficacy of split ME and split PE under their corresponding optimal condition. nCas9-RT split means nCas9 and RT are provided as two individual proteins, not a fused protein. For the split ME, the optimal condition producing the highest editing efficiency for RNA template and magRNA were 1000ng RNA template vector and 250 ng magRNA vector. The optimal condition of pegRNA producing the highest editing efficiency for the trans PE was the molar equivalent of the 1000ng RNA template expression vector. The molar ratio of the RNA template in PE system: pegRNA scaffold in PE: RNA template in ME: magRNA scaffold in ME was roughly equal to 1:1:1:0.25. The conditions of the common components, *i.e.*, MCP-RT (500ng), sp-nCas9 (500ng), and 2nd gRNA (250ng), were the same in between the ME and PE systems.

FIGs. 35 to 37 show results of assays that compare a PE3 system and its dual ME counterpart in editing the endogenous HEK3 site by installing 5 different types of mutations in HEK293T cells. Both editing efficacy and the insertion of the gRNA scaffold component were examined.

FIGs. 35A, 35B, and 35C show the components of the dual ME system and its PE3 counterpart. FIG. 35A shows ME- specific components, the separating modules of RNA template and magRNA. FIG. 35B shows PE- specific component, pegRNA, in which the gRNA and RT template are covalently linked. FIG. 35C shows the common components of ME and PE.

FIG. 36A shows the sequence of the HEK3 site and its complement (SEQ ID NOs: 10 and 11) with the annotation of gRNA guides. The RT templates in both ME and PE systems, T43-P13-5mut, contain 5 mutations across the template. For illustration purposes, only the 5' end and 3' end sequences of the target site are shown.

FIG. 36B shows representative Sanger sequencing chromatographs of the untreated (UT) and edited HEK3 genomic DNA (SEQ ID NO: 14). Arrows in the edited sample, 5mut, indicate the positions of the 5 edited bases.

FIG. 36C shows the quantification of editing efficacy detected by Sanger sequencing at five positions under each editing condition.

FIG. 36D shows results of assays that compared the editing efficiency between the PE3 system and its ME counterpart at 5G>T position. * $P < 0.05$, ** $P < 0.01$, $n = 3$.

FIGs. 37A, 37B, and 37C show results of assays that re-analyzed the representative samples from FIG. 36 by Next Generation Sequencing (NGS), and compared the ME and PE systems in terms of editing efficiency and the insertion caused by RT synthesis (reverse transcription) using the gRNA scaffold as a template. FIG. 37A shows editing efficacy at five positions by ME and PE editing. FIG. 37B shows the number of reads containing an insertion at the end of the template sequence with 3 or more base pairs identical to the gRNA scaffold 3' sequence. The number of edited reads and the ratio of insertion reads to edited reads are also shown in the table. ME-250, PE3-250, ME-500, PE3-500 were conditions in FIG. 37A, varying the amount of second gRNA (250ng vs 500 ng). FIG. 37C shows the ratios of reads containing gRNA scaffold-templated insertion to edited reads.

FIGs. 38A and 38B show results of assays that compare a dual ME system and its PE3 counterpart in editing the endogenous HEK3 site by deleting or inserting 3-nt sequences in HEK293T cells.

FIG. 38A shows the HEK3 site and its complement (SEQ ID NOs: 10 and 11) and edited sites with intended insertion (CTT) (SEQ ID NOs: 15-16) or deletion (GCA) (SEQ ID NOs: 17-18).

FIG. 38B shows editing efficacy by ME and PE3 in installing insertion and deletion analyzed by Sanger sequencing. * $P < 0.05$, $n = 3$.

FIG. 39 shows results of assays that compare the editing efficiency of a dual ME system and its counterpart PE in installing a single nucleotide insertion in episomal EGFP plasmids containing a 1-nt deletion at various positions in K562 cells. The $\Delta 4A$, $\Delta 20C$, and $\Delta 35C$ are positions relative to the EGFP nick site at G202 (FIG. 17C). The ME components were the same as in FIG. 25.

FIGs. 40A and 40B show results of assays that compare the editing efficiency of a dual ME system and its counterpart PE in creating a 29-nt deletion in a Dmd exon 23 skipping reporter plasmid in K562 cells. The ME and PE components were the same as in FIG. 28,

where the experiments were performed in HEK293T cells. FIG. 40A shows editing efficiency quantified by the percentage of GFP-positive cells, while FIG. 40B shows editing efficiency quantified by Sanger sequencing.

FIGs. 41A, 41B, and 41C show results of assays that compares a dual ME system and its PE3 counterpart in editing the endogenous HEK3 site by deleting or inserting a 3-nt sequence in K562 cells. Both editing efficacy and the insertion rate caused by RT template extension into the gRNA scaffold were compared. The ME components were the same as described in FIG. 38, where the experiment was performed in HEK293T cells. FIGs. 41A and 41B show the efficacy of CTT insertion and GCA deletion analyzed by Sanger sequencing and Next Generation Sequencing (NGS), respectively. FIG. 41C shows the ratios of NGS gRNA scaffold insertion reads to edited reads.

FIGs. 42A and 42B show the gene editing efficacy of a dual ME system in installing a point mutation at the endogenous HBB (hemoglobin beta) near the sickle cell anemia E6V mutation site in K562 cells. FIG. 42A shows the HBB sequence exon 1 and its complement (SEQ ID NOs: 19 and 20), with annotations of magRNA guide, second nick guide, nicking sites, as well as the intended +5G>T mutation (boxed within the first PAM) encoded in the RNA template. The adenine base immediately 5' to +5G base was mutated in sickle cell anemia E6V. FIG. 42B shows the efficacy of the ME in installing the +5G>T mutation in K562 cells.

FIGs. 43A, 43B, 43C, and 43D show results of assays that compare the editing efficacy of a saCas9 dual ME system and its counterpart saCas9 PE system. In this study, nickase saCas9 (N580A)-RT fusion and saCas9 gRNA scaffold were used in both systems. The efficiency in installing a 1-nt insertion and a silent point mutation in the EGFPΔ100G plasmid in HEK293 was compared.

FIG. 43A shows the EGFP Δ100G target sequence (5' tgagcaagggcgaggagctgttcaccggggtggtgcccacatcctggtcgagctggacggcg acgtaaacggccacaagttcagcgtgtccggcgaggcgagggcgatgccacctacggcaagc tgaccctgaagttcatctgcaccaccggcaag 3', SEQ ID NO: 21) and its complement (5' cttgccggtggtgcagatgaacttcagggtcagcttgccgtaggtggcatcgccctcgcc tcgccggacacgctgaacttgtggccggtttacgtcgccgtccagctcgaccaggatgggcac caccgccgtgaacagctcctcgcccttgetca, SEQ ID NO: 22), with magRNA guide, second nicking guide, saCas9 PAMs, as well as the intended editing sequence (5' tgagcaagggcgaggagctgttcaccggggtggtgcccacatcctggtcgagctggacggcg acgtaaacggccacaagttcagcgtgtccggcgagGgcgagggcgatgccacctacggcaag

ctgaccTtgaagttcatctgcaccaccggcaag 3' SEQ ID NO: 23) and its complement (5' cttgccggtggtgcagatgaacttcaAggtcagcttgccgtaggtggcatcgccctcgcCctcgccggacacgctgaacttgtggccgtttacgtcgccgtccagctcgaccaggatgggcacacccccggtgaacagctcctcgcccttgctca 3', SEQ ID NO: 24) annotated. For illustration purposes, only the 5' end, the deletion or insertion, and the 3' end sequences are shown.

FIG. 43B illustrates the conversion of non-fluorescent GFP to fluorescent GFP. FIGs. 43C-D compares editing efficiency by ME and PE, as quantified by GFP conversion (FIG. 43C) and Sanger sequencing (FIG. 43D, insertion).

DETAILED DESCRIPTION OF THE INVENTION

This disclosure relates to a novel RNA-guided gene editing system, called the Match Editing system or ME system, that utilizes a reverse transcriptase enzyme to copy a desired sequence information from an RNA template to DNA and insert the DNA into a target site in genome or extrachromosomal DNA. The gene editing system includes various novel features, such as (1) an editing RNA template that can be a fully independent RNA molecule and free of appendicular recruiting aptamer sequences such as MS2 aptamer, (2) the gRNA can be modified to contain a short polyribonucleotide tag that is complementary to a region within the template, as a result, (3) the template can be anchored to the editing location through Watson-Crick base pairing interaction between the editing RNA template and its matching gRNA. No prior design used a modified gRNA for recruiting an RNA template directly.

Differing in design and configuration from previously reported systems, the new system described herein provides an efficacious platform for convenient gene editing and *de novo* gene composing at target loci in genome or extrachromosomal DNA. The new design allows gene editing in target sequences farther from a PAM motif, is capable of editing all types of point mutations, and is proficient in insertion and deletion. Moreover, the modular feature of design allows cooperative multiplexing to enhance functionality. The new system and related platform therefore add a unique versatile tool to the gene editing technology.

As disclosed herein, an inventive feature of the ME system is that a gRNA can be engineered for direct RNA template recruitment through enabling its Watson-Crick sequence pairing with a natural template sequence. First of all, no prior arts taught about engineering gRNA sequence for direct recruitment of other gene editing components (*e.g.*, a RNA template) through Watson-Crick sequence complementation. Second, prior art in reverse transcriptase-mediated gene editing, including Prime Editing, did not teach about engineering gRNA

sequence for direct recruitment of an RNA template through Watson-Crick sequence complementation. The Match Editing feature is unexpected and not obvious over any prior art, neither RNA guided gene editing engineering in general nor reverse transcriptase-mediated gene editing such as Prime Editing.

Moreover, the distinct anchoring mechanism between the engineered gRNA, which is named magRNA in this disclosure, and the RNA template enable a number of new features that are not present in any other reverse transcriptase mediated gene editing systems.

First, the recruitment mechanism is through base-pairing between the natural template sequence and their complementary sequence added onto magRNA. No foreign or appendicular sequences, such as an MS2 aptamer, is needed in the RNA template. In the meantime, no trans-acting elements such as an MCP fusion is required for the template recruitment. This design ensures effective gene editing; in the meantime, the design is streamlined and compact compared to other similar reverse transcriptase mediated genome editing systems, an advantage for effectively delivery.

Second, the modular and versatile anchoring mechanism makes it convenient to engineer complex systems with multiplexing modules, either working in synergy to enhance efficacy or adding new functions. For example, two modules, each containing a magRNA with a tag matching a separate sequence within the same RNA template, could cooperatively recruit an RNA template hence drastically increase editing efficacy. In addition to the cooperative recruitment of RNA template by two separate modules, each module can also provide a distinct effector working synergistically together. For example, one module may contain an RNA aptamer to recruit a reverse transcriptase fused to an aptamer binding protein, the other module containing another RNA aptamer to recruit a chromatin modification enzyme fused to the corresponding aptamer binding protein. The physical and functional interaction between the two modules is re-enforced by interaction to the same RNA template recruited by the two magRNAs.

Third, the non-requirement for exogenous sequences in the RNA template mechanistically eliminates all source of “scars” currently affecting the reverse transcriptase mediated gene editing. One of the most challenging issues in introducing a gene or genetic fragment into a genome in a sequence-specific manner is the simultaneous introduction of unwanted exogenous sequences, or “scars.” Up to date, no effective “scar-free” solution is available when various enzymes such as reverse transcriptases, recombinases, and integrases are used to generate transgene insertion in non-germline mammalian cells. The underlying

cause of “scar” is that the template, either an RNA or DNA molecule, needs to contain a handle for the specific interaction and recruitment between the template molecule and genetic manipulation machinery, in addition to contain the genetic information to be inserted into the target site. Examples of these appendicular sequences for interaction with the genetic manipulation system are the flanking LTR, ITR, and RNA aptamers. These appendicular sequences are the sources of genetic “scars.” In particular relevant to this disclosure, it has been well-documented that Prime Editors often lead to insertion of unwanted pegRNA components, such as a part or the whole gRNA CRISPR interacting scaffold, into the target editing site, in addition to the desired RNA template within the pegRNA (Peter J. Chen *et al*, Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. Cell 184, 5635–5652, October 28, 2021. *e.g.* Figure 2 therein).

Fourth, the system design separates gRNA from the RNA template, and in the meantime installs a different template recruitment function to gRNA through a non-covalent Watson-Crick base-pairing mechanism. The combination makes it possible to utilize long RNA templates for effective gene editing without the intrinsic issue that the long RNA template may compromise the secondary structure of gRNA if the template and gRNA are covalently linked.

Fifth, structurally separating of the RNA template from the gRNA scaffold allows for varying ratio between RNA template and gRNA/nCas9/reverse transcriptase for optimal editing outcomes. In other words, in the pegRNA design, the ratio between the RNA template and gRNA molecule is always 1:1. As a result, at the editing site, one molecule of nCas9-RT is associated with one pegRNA (one gRNA and a covalently linked template). In contrast, with ME, one can vary the RNA template to gRNA ratio to increase editing efficacy and reduce non-specific editing. For example, instead of only 1:1 ratio can be achieved with pegRNA, a 1: 0.25 ratio can be implemented in ME, when desired.

As a result, by adopting the unique magRNA design for direct anchoring and recruiting a separate “scar-less” modular RNA template molecule, the ME systems disclosed herein overcome a number of existing critical obstacles (*e.g.*, intrinsic “scar” liability) in gene editing arena and in the meantime increase the flexibility and versatility (*e.g.*, by allowing for ratio change between the RNA template to gRNA and reverse transcriptase) for creating conditions for optimal genome editing (*e.g.*, high efficacy and low to negligible off-target effect). The ME thus is a unique system with unique features and improvements both in functionality and in elegance in design, as comparing to existing systems.

Match Editing (ME) System

As described herein, the ME system includes a number of unique features. For example, a desired genetic sequence can be scripted by an independent RNA template molecule that is separate from the gRNA AND free of protein recruiting motif such as an MS2 aptamer, or any other sequence not intended to be present in the final product. In other word, ME allows freedom to design a mechanistically scar-less template. Also, a separate matching gRNA molecule contains a short RNA sequence, or a matching tag, that is complementary to a region in the template RNA molecule. Further, the recruitment of the template is mediated by the Watson-Crick pairing between the template RNA and the matching sequence within the gRNA. The RNA template herein is often referred to as a Modular RNA template, as it is separate and independent from the sequence-specific DNA binding module. The gRNA with a matching tag complementary to a region within the template is called magRNA.

The separation of the DNA sequence recognition module (gRNA) from the genetic sequence information to be incorporated into the genome (RNA template) makes the system highly modular. As a result, a longer template can be designed independently without the concern of interfering the secondary structure and function of the gRNA. Moreover, the anchoring mechanism of Watson-Crick base pairing between RNA molecules, in contrast to affinity binding between an RNA motif (*e.g.*, MS2 RNA aptamer) and a protein (*e.g.*, aptamer binding protein MCP), eliminates unnecessary components such as the recruiting RNA aptamer sequence within the template and the aptamer binding fusion moiety in the CRISPR protein. This not only makes the system compact, but also eliminates source of editing “scars.” More importantly, the modular design of the system and simple base-pairing recruitment mechanism make it possible to build complex systems by simple and synergistical multiplexing the individual modules, linked (“clicked”) by nucleotide sequence complementation to various regions of the same RNA template.

In some embodiments, this disclosure provides examples of various basic single-module ME configurations as well as dual-module ME configurations by “clicking” two single module ME for achieving versatile gene editing and gene *de novo* composition functions.

Shown in FIG 1 is a single-module ME basic configuration, with a reverse transcriptase fused with an RNA guide nickase, using CRISPR/Cas9 protein as an example. This single-module Match Editing system uses a CRISPR/gRNA complex for RNA-guided sequence-recognition. The system comprises (a) a modular RNA template that does not contain any RNA recruiting elements such as RNA aptamers (*e.g.*, MS2 or PP7), (b) a magRNA containing a 5'

spacer sequence for target DNA recognition, an RNA scaffold for CRISPR/Cas complexing, and a 3' match RNA sequence that is complementary to a region in the modular template, and (c) a nickase CRISPR protein, nCRISPR/nCas9(H840A), fused with a reverse transcriptase (MMLV-RT) variant through a polypeptide linker. As illustrated in FIG. 1 the single module

basic system can comprise:

(1) an RNA template for reverse transcription (or a nucleotide encoding the expression of the template), containing a nucleotide sequence to be reverse transcribed to replace a desired sequence at a target site, with a 3' end sequence complementary to the nicking DNA strand at the target site;

(2) an engineered magRNA comprising

- a. a spacer sequence for sequence-specific recognition at the target site;
- b. an RNA scaffold for complexing with the RNA-guided nickase protein; and
- c. an RNA sequence tag (6-24 nucleotide in size) complementary to a sequence within the RNA template (1); and

(3) an RNA guided nickase (*e.g.*, nCRISPR/nCas9 H840A) fused with a reverse transcriptase via a polypeptide linker.

While as illustrated in FIG. 1 the matching sequence is at 3' of the gRNA, the tag can be localized at other locations within the gRNA, such as at the stem-loop region or at tetraloop region, or 5' of the gRNA.

FIG. 2 illustrate how the Match Editing system works. As illustrated in FIG. 2, the Match Editing system recognizes a DNA target site through the spacer sequence at the 5' end of the engineered gRNA, forming a R-loop at the target DNA locus. The tag sequence at the 3' end of magRNA binds to a segment within the RNA template that is complementary to the tag, anchoring the template to the site. In the meantime, the nickase activity leads to a single strand DNA break at the non-target strand (*e.g.*, 3 nucleotide upstream of the PAM motif if an nCas9 H840A nickase is used). The 3' end of the RNA template then anneals with the nicked complementary non-target DNA strand in the R-loop. The non-target DNA nicking end then serves as a primer for the reverse transcriptase, using the RNA template. As a result, the sequence information in the modular RNA template is copied to a DNA strand, which is subsequently incorporated into the target DNA molecule.

FIG. 3 shows a cellular repair mechanism that leads to incorporation of the newly synthesized DNA strand into a target DNA locus. More specifically, FIG. 3 illustrates the

process how Match Editing leads to precision genome editing. The process includes the following:

- A. A nicking is introduced by nickase at the non-target strand within the target site;
- B. Reverse transcriptase copies the RNA template sequence into a DNA sequence primed by the nicking DNA strand;
- C. the newly synthesized 3'-DNA flap competes the 5'-DNA flap;
- D. the 5'-flap is preferentially removed in the cell by nucleases, the newly synthesized DNA strand anneals to the opposite strand; and
- E. DNA repair either removes the unmatched sequences in the "old" strand, leading to editing, or removes the unmatched sequence in the newly synthesized strand, leading to a no editing event.

The newly synthesized strand may carry desired point mutations (either transversions or transitions or both), insertions, or deletions.

Shown in FIG. 4 is one special example of magRNA-RNA template interaction. In this example, the 3'-matching tag sequence of magRNA is complementary specifically to the 5'-end sequence of RNA template.

Shown in FIG. 5 is another special example with a magRNA having a 5' match tag. In this example, the tag that is complementary to a region of RNA template is located at the 5' end of gRNA. A linker polynucleotide sequence is located between the 5' match tag sequence and the spacer sequence.

FIG. 6 shows an exemplary variant of Match Editing system where the reverse transcriptase is provided *in split* rather than directly fused to the RNA-guided sequence-specific nickase. As illustrated in FIG. 6, the RT *in split* system comprises the following:

- (1) an RNA template for reverse transcription, containing a polynucleotide sequence to be reverse transcribed to replace a desired sequence at a target site, and a 3' end sequence complementary to the nicking strand at the target site;
- (2) an RNA-guided nickase for sequence recognition;
- (3) an engineered magRNA comprising
 - a. a 5' guide for sequence-specific recognition at the target site
 - b. an RNA scaffold for complexing with the RNA-guided nickase protein
 - c. an RNA aptamer sequence at the loops of the RNA scaffold, such as stem loop or tetra loop

- d. a 3' extension sequence (6-24 nucleotide in size) complementary to a sequence segment locating at the 5' end or in the middle of the RNA template (1); and

(4) a protein comprising a reverse transcriptase fusing with the cognate protein (*e.g.*, MCP) binding to the aptamer within the gRNA scaffold (*e.g.*, MS2) through a polypeptide linker.

FIG. 7 is an illustration showing how the RT *in split* system works. As illustrated in FIG. 7, the RT *in split* system works in way that is very similar to the schematic as illustrated in FIG. 2, except that the reverse transcriptase (RT) is not fused to the RNA-guided sequence-specific complex protein, *e.g.*, nCas9. Instead, the gRNA containing a MS2 aptamer at the stem loop of the gRNA (or other location), which recruits the fusion protein of MCP-RT.

FIG. 8 shows a variant configuration of RT *in split* system, where the 3' magRNA extension sequence complementary to the sequence 5' end of RT template.

FIG. 9 shows two Dual ME systems each having a nicking module that creates a single strand DNA break at a close-by downstream location at the opposite strand. Shown in FIG. 9A is a Dual ME system with fusing RT and magRNA-gRNA pairing. The second module is for creating a second nick for increasing editing efficiency. Shown in FIG. 9B is a Dual ME system with split RT and magRNA with MS2-gRNA 0xMS2. The second module is for creating a second nick for increasing editing efficiency. Each dual ME system increases editing efficiency through mechanisms illustrated in FIG. 10.

More specifically, as shown in FIG. 10, mechanistically the second nicking favors the removal of the unmatched sequence in the nicking strand, leading to higher editing efficiency. As a result, a second nicking at the nearby opposite strand mediated by dual ME greatly enhance target editing efficiency.

Disclosed herein is a Single Template-Coupled Dual ME system for editing a target site using a long RNA template. Two examples, Coupled Dual ME with fused RT (FIG. 11A) and Coupled Dual ME with RT *in split* (FIG. 11B), are illustrated. Each module of the system contains a magRNA with a match tag complementary to a separate region of the RNA template. Anchoring of the RNA template with two magRNA strengthens the interaction between the editing complexes and the template, hence increasing editing efficiency or enhancing the functionality of individual module.

FIG. 11 also illustrated the principle of how two (or more) individual ME modules are physically and functionally linked, or “clicked”, together for enhancing activity or creating new

synergistic functionality, through simple base-pairing complementation to different regions of the same RNA template.

Also disclosed herein is a Coupled Dual ME for enhancing the functionality. Several examples (a RT fusion version and a RT- *in split* version) are illustrated in FIG. 12 and their mechanisms are explained in FIG. 13. In these examples, the individual module is designed to function cooperatively. For example, in the RT fusion version, the 3' end of the RNA template is designed as identical to the 3' of the second nicking strand. As a result, the second nicking strand serves as the primer for synthesizing the second DNA strand using the newly synthesized DNA strand as a template.

FIG 13 illustrates the scheme of cooperative interaction involving a long RNA template and synthesis of the second strand DNA by the single-template coupled Dual ME system: (A)-(B) first nicking and reverse transcription of the long template by the first module; (C)-(D) second nicking by the second module; (E)-(F) second DNA strand synthesis mediated by the DNA-dependent DNA polymerase activity of a reverse transcriptase, facilitated by the complementation between the 3' end of the second nicking site and the 5' end of the RNA template; (G)-(H) removal of 5' flap.

Also disclosed herein is an example system that can be used for insertion or deletion. As illustrated in FIG. 14, Module 1 and Module 2 are separate ME module each having separate components as described before, including separate RNA template. In this configuration, although the two modules do not physically interact with each other through a common RNA template, they can be designed to cooperate functionally. For example, the RNA templates can be designed in such a way that the reverse transcribed DNA strand partially complementary with each other.

The illustrated two modules create a nick *in trans* in opposite strands. Two modules or multiplex modules can be functional in both *cis* (with PAM and guide sequence on the same strand) or *in trans*, to achieve various functionalities, such as tandemly insertion small fragments to achieve insertion of a large fragment, or correcting multiple mutations of various defects.

FIG. 15 is a schematic showing on how the dual-template dual-module ME system, can lead to *de novo* gene composition (insertion) or large fragment gene deletion. As illustrated in FIGs. 15A and B, module 1 ME and module 2 ME mediate nicking at their corresponding target sites and lead to first strand DNA synthesis by reverse transcription of their corresponding RNA template. This then leads to annealing of the 3' ends of the two first DNA

strands through the complementary sequences as shown in FIG. 15C. This further allows second strand DNA synthesis by DNA-dependent DNA polymerase activity of the reverse transcriptase or endogenous DNA polymerase activity as shown in FIG. 15D. After removal of 5' flaps and ligation (FIG. 15E), an insertion or deletion is achieved (FIG. 15F). The mechanism can lead to both insertion and deletion (*de novo* gene composition) depending on the content and length of the two RNA templates and the sequences removed.

Various second effectors can be included in any of the systems described herein. They can be provided *in split* or by fusion as shown in FIG. 16. Examples can include proteins that facilitate the editing process or increase editing efficiency, such as chromatin modification enzymes, inhibitors of mismatch repair enzymes, and proteins that can stabilize template RNA and magRNA. The examples of the effectors are a human RNase Inhibitor protein, a dominant negative protein of MMR pathway protein including MLH1 protein, a 5' DNA nuclease Fen1 protein, or the like that could enhance the editing efficiency. The module illustrated in FIG. 16 is intended to "clicking" or multiplexing into the dual or complex systems exemplified in FIG. 9 to FIG. 15.

RNA-guided nickase

A key second component of the gene editing complex or system disclosed herein is an RNA-guided nickase. A nickase is an enzyme that creates a single-strand break (also known as a "nick") in double strand DNA, *i.e.*, cuts one strand but not the other of the DNA double helix.

As used herein, an "RNA-guided nickase" means a polypeptide or complex of polypeptides having DNA nickase activity, wherein the DNA nickase activity can be sequence-specific and depends on the sequence of the RNA. Exemplary RNA-guided nickases include Cas nickases. Cas nickases include, but are not limited to, nickase forms of a Csm or Cmr complex of a type III CRISPR system, the Cas10, Csm1, or Cmr2 subunit thereof, a Cascade complex of a type I CRISPR system, the Cas3 subunit thereof, and Class 2 Cas nucleases. Class 2 Cas nickases include Class 2 Cas nuclease variants in which only one of the two catalytic domains is inactivated, which have RNA-guided DNA nickase activity. Class 2 Cas nickases include, for example, Cas9 (*e.g.*, H840A, D10A, or N863A variants of SpyCas9), Cpf1, C2c1, C2c2, C2c3, HF Cas9 (*e.g.*, N497A, R661A, Q695A, Q926A variants), HypaCas9 (*e.g.*, N692A, M694A, Q695A, H698A variants), eSPCas9(1.0) (*e.g.*, K810A, KI 003 A, R1060A variants), and eSPCas9(1.1) (*e.g.*, K848A, K1003A, R1060A variants) proteins and modifications thereof. Cpf1 protein, Zetsche *et al.*, Cell, 163: 1-13 (2015), is homologous to Cas9, and contains a RuvC-like protein domain. Cpf1 sequences of Zetsche are incorporated by

reference in their entirety. See, *e.g.*, Zetsche, Tables SI and S3. “Cas9” encompasses 5. *S. pyogenes* (Spy) Cas9, the variants of Cas9 listed herein, and equivalents thereof. See, *e.g.*, Makarova *et al.*, Nat Rev Microbiol, 13(11): 722-36 (2015); Shmakov *et al.*, Molecular Cell, 60:385-397 (2015). Makarova *et al.*, NAT. REV. MICROBIOL, 18:67-83 (2020).

5 In some embodiments, an RNA-guided nickase disclosed herein is a Cas nickase. In some embodiments, an RNA-guided nickase is from a specific Cas nuclease with its catalytic domain(s) being inactivated. In some embodiments, the RNA-guided nickase is a Class 2 Cas nickase, such as a Type II Cas9 nickase or a Cpf1 nickase. In some embodiments, the RNA-guided nickase is an *S. pyogenes* Cas9 nickase. In some embodiments, the RNA-guided nickase is *Neisseria meningitidis* Cas9 nickase. In some embodiments, the RNA-guided nickase is *Staphylococcus aureus* Cas9 nickase.

10 In some embodiments, the Class 2 Cas is a Type V Cas protein, such as a Cas12. In some embodiments, the Cas12 is Cpf1 (Cas12a), C2c1 (Cas12b), C2c3 (Cas12c), CasY (Cas12d), CasX (Cas12e), Cas14 (Cas12f), CasPhi (Cas12j), and the orthologs and variants thereof.

15 In some embodiments, the Cas protein is Type VI Cas protein, such as a Cas 13. In some embodiments, the Cas13 is Cas13a, Cas13b, Cas13c, Cas13d, Cas13x, and the orthologs and variants thereof.

20 In some embodiments, an RNA-guided nickase disclosed herein is a variant of transposon-encoded IscB, IsrB, or TnpB family endonuclease protein.

In some embodiments, the RNA-guided nickase is a modified Class 2 Cas protein or derived from a Class 2 Cas protein. In some embodiments, the RNA-guided nickase is modified or derived from a Cas protein, such as a Class 2 Cas nuclease (which may be, *e.g.*, a Cas nuclease of Type II, V, or VI). Class 2 Cas nuclease include, for example, Cas9, Cpf1, C2c1, C2c2, and C2c3 proteins and modifications thereof. Examples of Cas9 nucleases include those of the type II CRISPR systems of *S. pyogenes*, *S. aureus*, and other prokaryotes (see, *e.g.*, the list in the next paragraph), and modified (*e.g.*, engineered or mutant) versions thereof. See, *e.g.*, US2016/0312198 A1 and US 2016/0312199 A1, which are incorporated by reference in their entireties. Other examples of Cas nucleases include a Csm or Cmr complex of a type III CRISPR system or the Cas 10, Csm1, or Cmr2 subunit thereof; and a Cascade complex of a type I CRISPR system, or the Cas3 subunit thereof. In some embodiments, the Cas nuclease may be from a Type-IIA, Type-IIB, or Type-IIC system. For discussion of various CRISPR systems and Cas nucleases, see, *e.g.*, Makarova *et al.*, NAT. REV. MICROBIOL. 9:467-477

(2011); Makarova *et al.*, NAT. REV. MICROBIOL, 13: 722-36 (2015); Shmakov *et al.*, MOLECULAR CELL, 60:385-397 (2015). Makarova *et al.*, NAT. REV. MICROBIOL, 18:67-83 (2020).

A Cas nickase described herein may be a nickase form of a Cas nuclease from the species including, but not limited to, *Streptococcus pyogenes*, *Streptococcus thermophilus*, *Streptococcus sp.*, *Staphylococcus aureus*, *Listeria innocua*, *Lactobacillus gasseri*, *Francisella novicida*, *Wolinella succinogenes*, *Sutterella wadsworthensis*, *Gammaproteobacterium*, *Neisseria meningitidis*, *Campylobacter jejuni*, *Pasteurella multocida*, *Fibrobacter succinogene*, *Rhodospirillum rubrum*, *Nocardiopsis dassonvillei*, *Streptomyces pristinaespiralis*, *Streptomyces viridochromogenes*, *Streptomyces viridochromogenes*, *Streptosporangium roseum*, *Streptosporangium roseum*, *Alicyclobacillus acidocaldarius*, *Bacillus pseudomyoides*, *Bacillus selenitireducens*, *Exiguobacterium sibiricum*, *Lactobacillus delbrueckii*, *Lactobacillus salivarius*, *Lactobacillus buchneri*, *Treponema denticola*, *Microscilla marina*, *Burkholderiales bacterium*, *Polaromonas naphthalenivorans*, *Polaromonas sp.*, *Crocospaera watsonii*, *Cyanothece sp.*, *Microcystis aeruginosa*, *Synechococcus sp.*, *Acetohalobium arabaticum*, *Ammonifex degensii*, *Caldicelulosiruptor beccsii*, *Candidates Desulforudis*, *Clostridium botulinum*, *Clostridium difficile*, *Finegoldia magna*, *Natranaerobius thermophilus*, *Pelotomaculum thermopropionicum*, *Acidithiobacillus caldus*, *Acidithiobacillus ferrooxidans*, *Allochromatium vinosum*, *Marinobacter sp.*, *Nitrosococcus halophilus*, *Nitrosococcus watsoni*, *Pseudoalteromonas haloplanktis*, *Ktedonobacter racemifer*, *Methanohalobium evestigatum*, *Anabaena variabilis*, *Nodularia spumigena*, *Nostoc sp.*, *Arthrospira maxima*, *Arthrospira platensis*, *Arthrospira sp.*, *Lyngbya sp.*, *Microcoleus chthonoplastes*, *Oscillatoria sp.*, *Petrogona mobilis*, *Thermosiphon africanus*, *Streptococcus pasteurianus*, *Neisseria cinerea*, *Campylobacter lari*, *Parvibaculum lavamentivorans*, *Corynebacterium diphtheria*, *Acidaminococcus sp.*, *Lachnospiraceae bacterium ND2006*, or *Acaryochloris marina*.

In some embodiments, the Cas nickase is a nickase form of the Cas9 nuclease from *Streptococcus pyogenes*. In some embodiments, the Cas nickase is a nickase form of the Cas9 nuclease from *Streptococcus thermophilus*. In some embodiments, the Cas nickase is a nickase form of the Cas9 nuclease from *Neisseria meningitidis*. See *e.g.*, WO/2020081568, describing an Nme2Cas9 D16A nickase. In some embodiments, the Cas nickase is a nickase form of the Cas9 nuclease from *Staphylococcus aureus*. In some embodiments, the Cas nickase is a nickase form of the CpfI nuclease from *Francisella novicida*. In some

embodiments, the Cas nickase is a nickase form of the Cpf1 nuclease from *Acidaminococcus* sp. In some embodiments, the Cas nickase is a nickase form of the Cpf1 nuclease from *Lachnospiraceae* bacterium ND2006. In further embodiments, the Cas nickase is a nickase form of the Cpf1 nuclease from *Francisella tularensis*, *Lachnospiraceae* bacterium, *Butyrivibrio proteoclasticus*, *Peregrinibacteria bacterium*, *Parcubacteria bacterium*, *Smithella*, *Acidaminococcus*, *Candidatus Methanoplasma termitum*, *Eubacterium eligens*, *Moraxella bovoculi*, *Leptospira inadai*, *Porphyromonas crevioricanis*, *Prevotella disiens*, or *Porphyromonas macacae*. In certain embodiments, the Cas nickase is a nickase form of a Cpf1 nuclease from an *Act* *aminococcus* or *Lachnospiraceae*. As discussed elsewhere, a nickase may be derived from (*i.e.* related to) a specific Cas nuclease in that the nickase is a form of the nuclease in which one of its two catalytic domains is inactivated, *e.g.*, by mutating an active site residue essential for nucleolysis, such as D10, H840, or N863 in Spy Cas9. One skilled in the art will be familiar with techniques for easily identifying corresponding residues in other Cas proteins, such as sequence alignment and structural alignment, which is discussed in detail below.

In other embodiments, the Cas nickase may relate to a Type-I CRISPR/Cas system. In some embodiments, the Cas nickase may be a component of the Cascade complex of a Type-I CRISPR/Cas system. In some embodiments, the Cas nickase may be a Cas3 protein. In some embodiments, the Cas nickase may be from a Type-III CRISPR/Cas system.

In some embodiments, a Cas nickase is a nickase form of a Cas nuclease or a modified Cas nuclease in which an endonucleolytic active site is inactivated, *e.g.*, by one or more alterations (*e.g.*, point mutations) in a catalytic domain. See, *e.g.*, US Pat. No. 8,889,356 for discussion of Cas nickases and exemplary catalytic domain alterations.

Wild type *S. pyogenes* Cas9 has two catalytic domains: RuvC and HNH. The RuvC domain cleaves the non-target DNA strand, and the HNH domain cleaves the target strand of DNA. In some embodiments, a Cas nuclease may comprise an amino acid substitution in the RuvC or RuvC-like nuclease domain. Exemplary amino acid substitutions in the RuvC or RuvC-like nuclease domain include D10A (based on the *S. pyogenes* Cas9 protein). See, *e.g.*, Zetsche *et al.* (2015) Cell Oct 22:163(3): 759-771. In some embodiments, the Cas nuclease may comprise an amino acid substitution in the HNH or HNH-like nuclease domain. Exemplary amino acid substitutions in the HNH or HNH-like nuclease domain include E762A, H840A, N863A, H983A, and D986A (based on the *S. pyogenes* Cas9 protein). See, *e.g.*, Zetsche *et al.* (2015). Further exemplary amino acid substitutions include D917A, E1006A,

and D1255A (based on the *Francisella novicida* U112 CpfI (FnCpfI) sequence (UniProtKB - A0Q7Q2 (CPF1 FRATN))).

In some embodiments, a Cas nickase such as a Cas9 nickase has an inactivated RuvC or HNH domain. In some embodiments, a nickase is used having a RuvC domain with reduced activity. In some embodiments, a nickase is used having an inactive RuvC domain. In some
5 embodiments, a nickase is used having an HNH domain with reduced activity. In some
embodiments, a nickase is used having an inactive HNH domain. In some embodiments, a
Cas9 nickase has an active HNH nuclease domain and is able to cleave the non-targeted strand
of DNA, *i.e.*, the strand bound by the gRNA and has an inactive RuvC nuclease domain and is
10 not able to cleave the targeted strand of the DNA, *i.e.*, the strand where base editing by
deaminase is desired.

An exemplary Cas9 nickase amino acid sequence is provided below:

spCas9 (H840A) (SEQ ID NO:25)

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEA
15 TRLKRTARRRYTRRKNRICYLQEFSNEMAKVDDSFHRLVESFLVEEDKKHERHPIFGNIV
DEVAYHEKYPTIYHLRKKLVDDSTDKADLRILIYALAHMIKFRGHFLIEGDLNPDNSDVKL
FIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLG
LTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILSDILRV
NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQ
20 EEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPF
LKDNREKIEKILTFRIPIYVGPLARGNSRFAMTRKSEETITPWNFEVVDKGASAQSFIER
MTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK
TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDIL
EDIVLTLTFLFEDREMIEERLKTIAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGK
25 TILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQ
TVKVVDDELVKVMGRHKPENIVIEAMARENQTTQKGQKNSRERMKRIEIGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVAIVPQSFLKDDSIDNKVLTRS
DKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINN
30 YHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFY
SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGKSKKLKSVK
ELLGITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLEFLENRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADAN
35 LDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLI
HQSITGLYETRIDLSQLGGD

saCas9(N580A) (SEQ ID NO: 26)

MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRR
40 HRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEFSAALLHLAKRRGVHNVN
EVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSYVKEAKQL
LKVQKAYHQLDQSFIDTYIDLLETRRTYIEGPGEGSPFGWKDIKEWYEMLMGHCTYFPPE
LRSVKYAYNADLYNALNDLNNLVITRDENEKLEYEKFQIIENVFKQKKKPTLKQIAKEIL
VNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIEAELLDDQIAKILTIYQSSEDIQEELT

NLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQ
 QKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKR
 NRQTNERIEEIIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNPNFYEVVDHIIPR
 SVSFDNSFNKVLVKQEEASKKGNRTPFQYLSSSDSKISYETFKKHILNLAQKGRISKTKK
 5 EYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFL
 RRKWKFKKERNKGYKHHAEADALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMP
 EIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNL
 NGLYDKDNDKLLKLINKSPEKLLMYHHDPTQYQKLKLIMEQYGDEKNPLYKYEEETGNY
 LTKYSKKDNGPVIKKIKYYGNKLNALHDITDDYPNSRNKVVKLSLKPYPYFDVYLDNGVYK
 10 FVTVKNLDVIKKENYEVNSKCYEEAKKLKKISNQAEFASFYNNDLIKINGELRYVIGVNN
 DLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQII
 KK

Other examples are nickase version of the following Cas variants: SpCas9 D1135E
 variant, SpCas9 VRER variant, SpCas9 EQR variant, SpCas9 VQR variant, SpCas9-NG,
 15 xCas9, SpCas9-NG; *Staphylococcus aureus* (SA); SaCas9, *Acidaminococcus* sp. (AsCpf1) and
Lachnospiraceae bacterium (LbCpf1); AsCpf1 RR variant, LbCpf1 RR variant, AsCpf1 RVR
 variant, *Campylobacter jejuni* (CJ) Cas9, *Neisseria meningitidis* (NM) Cas9, *Streptococcus*
thermophilus (ST) Cas9, *Treponema denticola* (TD) Cas9

In some embodiments, the RNA-guided nickase comprises an amino acid sequence
 20 having at least 80%, 90%, 95%, 98%, or 99% identity to above sequence.

Reverse transcriptase

The gene editing complex or system disclosed herein includes a reverse transcriptase.
 As used herein, a "reverse transcriptase" describes a class of polymerases characterized as
 RNA-dependent DNA polymerases. All known reverse transcriptases require a primer to
 25 synthesize a DNA transcript from an RNA template. Historically, reverse transcriptase has
 been used primarily to transcribe mRNA into cDNA which can then be cloned into a vector for
 further manipulation.

Avian myoblastosis virus (AMV) reverse transcriptase was the first widely used RNA-
 dependent DNA polymerase (Verma, Biochim. Biophys. Acta 473:1 (1977)). The enzyme has
 30 5'-3' RNA-directed DNA polymerase activity, 5'-3' DNA-directed DNA polymerase activity,
 and RNase H activity. RNase H is a processive 5' and 3' ribonuclease specific for the RNA
 strand for RNA-DNA hybrids (Perbal, A Practical Guide to Molecular Cloning, New York:
 Wiley & Sons (1984)). A detailed study of the activity of AMV reverse transcriptase and its
 associated RNase H activity has been presented by Berger *et al.*, Biochemistry 22:2365-2372
 35 (1983). Another reverse transcriptase which is used extensively in molecular biology is reverse
 transcriptase originating from Moloney murine leukemia virus (M-MLV). See, *e.g.*, Gerard, G.
 R., DNA 5:271-279 (1986) and Kotewicz, M. L., *et al.*, Gene 35:249-258 (1985). M-MLV

reverse transcriptase substantially lacking in RNase H activity has also been described. See, *e.g.*, U.S. Pat. No. 5,244,797. The disclosure contemplates the use of any such reverse transcriptases, or variants or mutants thereof.

Reverse transcriptases are also from non-viral origin, including retrotransposons. Examples include reverse transcriptases from LTR retrotransposons, Endogenous retrovirus, Non-LTR retrotransposons, LINEs. Example proteins include Line 1 ORF2, R2Bm, R2OI.

The disclosure contemplates any wildtype reverse transcriptase obtained from any naturally-occurring organism or virus, or obtained from a commercial or non-commercial source. In addition, the reverse transcriptases usable in the gene editing complex or system of the disclosure can include any naturally-occurring mutant RT, engineered mutant RT, or other variant RT, including truncated variants that retain function. The RTs may also be engineered to contain specific amino acid substitutions, such as those specifically disclosed herein.

Reverse transcriptases are multi-functional enzymes typically with three enzymatic activities including RNA- and DNA-dependent DNA polymerization activity, and an RNaseH activity that catalyzes the cleavage of RNA in RNA-DNA hybrids. Some mutants of reverse transcriptases have disabled the RNaseH moiety to prevent unintended damage to the RNA. These enzymes that synthesize complementary DNA (cDNA) using mRNA as a template were first identified in RNA viruses. Subsequently, reverse transcriptases were isolated and purified directly from virus particles, cells or tissues. (*e.g.*, see Kacian *et al.*, 1971, Biochim. Biophys. Acta 46: 365-83; Yang *et al.*, 1972, Biochem. Biophys. Res. Comm. 47: 505-11; Gerard *et al.*, 1975, J. Virol.15: 785-97; Liu *et al.*, 1977, Arch. Virol.55:187- 200; Kato *et al.*, 1984, J. Virol. Methods 9: 325-39; Luke *et al.*, 1990, Biochem.29: 1764-69 and Le Grice *et al.*, 1991, J. Virol.65: 7004-07, each of which are incorporated by reference). More recently, mutants and fusion proteins have been created in the quest for improved properties such as thermostability, fidelity and activity. Any of the wild type, variant, and/or mutant forms of reverse transcriptase which are known in the art or which can be made using methods known in the art are contemplated herein.

Examples of sources for reverse transcriptases include, but are not limited to, Moloney murine leukemia virus (M-MLV or MLVRT); human T-cell leukemia virus type 1 (HTLV-1); bovine leukemia virus (BLV); Rous Sarcoma Virus (RSV); human immunodeficiency virus (HIV); yeast, including *Saccharomyces*, *Neurospora*, *Drosophila*; primates; and rodents. See, for example, Weiss, *et al.*, U.S. Pat. No. 4,663,290 (1987); Gerard, G. R., DNA:271-79 (1986); Kotewicz, M. L., *et al.*, Gene 35:249- 58 (1985); Tanese, N., *et al.*, Proc. Natl. Acad. Sci.

(USA):4944-48 (1985); Roth, M. J., *et al.*, J. Biol. Chem.260:9326-35 (1985); Michel, F., *et al.*, Nature 316:641-43 (1985); Akins, R. A., *et al.*, Cell 47:505-16 (1986), EMBO J.4:1267-75 (1985); and Fawcett, D. F., Cell 47:1007-15 (1986) (each of which are incorporated herein by reference in their entireties).

Exemplary enzymes can include, but are not limited to, M-MLV reverse transcriptase and RSV reverse transcriptase. Enzymes having reverse transcriptase activity are commercially available. In certain embodiments, the reverse transcriptase provided in trans to the other components of the ME system. That is, the reverse transcriptase is expressed or otherwise provided as an individual component.

A person of ordinary skill in the art will recognize that wild type reverse transcriptases, including but not limited to, Moloney Murine Leukemia Virus (M-MLV); Human Immunodeficiency Virus (HIV) reverse transcriptase and avian Sarcoma-Leukosis Virus (ASLV) reverse transcriptase, which includes but is not limited to Rous Sarcoma Virus (RSV) reverse transcriptase, Avian Myeloblastosis Virus (AMV) reverse transcriptase, Avian Erythroblastosis Virus (AEV) Helper Virus MCAV reverse transcriptase, Avian Myelocytomatosis Virus MC29 Helper Virus MCAV reverse transcriptase, Avian Reticuloendotheliosis Virus (REV-T) Helper Virus REV-A reverse transcriptase, Avian Sarcoma Virus UR2 Helper Virus UR2AV reverse transcriptase, Avian Sarcoma Virus Y73 Helper Virus YAV reverse transcriptase, Rous Associated Virus (RAV) reverse transcriptase, and Myeloblastosis Associated Virus (MAV) reverse transcriptase may be suitably used in the subject methods and composition described herein. Additional examples are described in Martín-Alonso, S. *et al.*, Trends in Biotechnology 39 (2): 194-210 (2021) and WO2020191248, the content of which is incorporated herein.

Example reverse transcriptase variant: MMLV-RT variant (D200N/T306K/W313F/T330P/L603W) (SEQ ID NO: 27).

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLI I PLKATSTPVS IK
QYPMSQEARLG I KPHIQRL LDQG I LVPCQSPWNTPLLPVKKPGTNDYRPVQDLREV NKRVED
IHPTVPNPYNLLSGLPPSHQWYTVLDLKD AFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWT
RLPQGFKNSPTLFNEALHRDLADFRIQH PDL I LLQYVDDLLLAATSELD CQQGTRALLQTLG
NLGYRASAKKAQICQKQVKYLG YLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFC
RLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDE
KQGYAKGVLTQKLGPPWRPVAYLSKKLD PVAAGWP PCLRMVAAIAVLTKDAGKLTMGQPLVI
LAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGP VVALNPATLLPLPEEGLQH NCL
DILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSA
QRAELIALTQALKMAEGKKLN VYTD SRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILAL
LKALFLPKRLS I IHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP

Nuclear Localization Signals

In some embodiments, the ME component proteins contain one or more nuclear localization signal (NLS) peptides. The NLS peptides can localize at the N-terminus or C-terminus or internal region of the protein. NLS peptides are generally short peptides that act as a signal fragment that mediates the transport of proteins from the cytoplasm into the nucleus. A list of commonly used NLS peptides can be found in Lu, J., Wu, T., Zhang, B. *et al.* Cell Commun Signal 19, 60 (2021). A full list of NLS is manually curated in the database genome.unmc.edu/LocSigDB/.

In some embodiments, the NLS is classic monopartite or bipartite nuclear localization signals (cNLS). The monopartite NLS is composed of 4-8 basic amino acids, generally contains four or more positively charged residues arginine (R) lysine (K). The bipartite NLS contains two clusters of basic amino acids, separated by a spacer of about 10 amino acids. In some embodiment, the bipartite NLS is the nucleoplasmin NLS peptide KR[PAATKKAGQA]KKKK (SEQ ID NO: 28).

In some embodiments, the monopartite NLS is SV40 Large T-antigen NLS peptide Pro-Lys-Lys-Lys-Arg-Lys-Val (PKKKRKV, SEQ ID NO: 29). In some embodiments, the SV40 NLS is at the N-terminus of the ME component proteins. In some embodiments, the SV40 NLS is at the C-terminus of the ME component proteins.

Linkers

In some embodiments, the reverse transcriptase and the RNA-guided nickase described herein is joined via a linker, such as, but not limited to chemical modification, peptide linkers, chemical linkers, covalent or non-covalent bonds, or protein fusion or by any means known to one skilled in the art. The joining can be permanent or reversible. See for example U.S. Pat. Nos. 4625014, 5057301 and 5514363, US Application Nos. 20150182596 and 20100063258, and WO2012142515, the contents of which are incorporated herein in their entirety by reference. In some embodiments, several linkers can be included in order to take advantage of desired properties of each linker and each protein domain in the conjugate. For example, flexible linkers and linkers that increase the solubility of the conjugates are contemplated for use alone or with other linkers. Peptide linkers can be linked by expressing DNA encoding the linker to one or more protein domains in the conjugate. Linkers can be acid cleavable, photocleavable and heat sensitive linkers. Methods for conjugation are well known by persons skilled in the art and are encompassed for use in the present disclosure.

In some embodiments, the linker can be an organic molecule, polymer, or chemical moiety. In some embodiments, the linker can be a peptide linker. In some embodiments, the peptide linker can be any stretch of amino acids having at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 40, at least 50, or more amino acids.

In some embodiments, the peptide linker is the 16 residue “XTEN” linker, or a variant thereof (See, e.g., Schellenberger *et al.* A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. Nat. Biotechnol. 27, 1186-1190 (2009)).

In some embodiments, the peptide linker comprises a (GGGGS)_n (SEQ ID NO: 30), a (G)_n, an (EAAAK)_n (SEQ ID NO: 31), a (GGS)_n, an SGSETPGTSESATPES motif (SEQ ID NO: 32) (see, e.g., Guilinger J P, Thompson D B, Liu D R. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nat. Biotechnol. 2014; 32(6): 577-82; the entire contents are incorporated herein by reference), or an (XP)_n motif, or a combination of any of these, wherein n is independently an integer between 1 and 30. See, WO2015089406 the entire content of which is incorporated herein by reference.

Matching gRNA (magRNA) molecule

Another key component of the gene editing complex or system disclosed herein is an RNA molecule called matching gRNA or magRNA. This magRNA may contain three sub-components: (1) a programmable guide RNA sequence or spacer sequence for target DNA recognition, (2) an RNA scaffold capable of binding to an RNA-guided nickase, and (3) an anchoring tag sequence complementary to a segment in an RNA template. This magRNA can be either a single RNA molecule or a complex of multiple RNA molecules.

As disclosed in some embodiments, the magRNA and the RNA-guided nickase (e.g., Cas protein) together form a complex (e.g., a CRISPR/Cas-based module) for sequence targeting and recognition. While the tag sequence brings a template to a target site to be edited, the reverse transcriptase can be recruited to the site by different means. In some examples, the reverse transcriptase can be recruited directly via fusion to the nickase. In others, the reverse transcriptase can be recruited *in split (individually)* by a recruiting RNA motif, which via an RNA-protein binding pair, recruits the reverse transcriptase.

In some embodiments, the match tag sequence in gRNA does not contain polyN wherein N is any repeating base of ribonucleotide or deoxyribonucleotide.

Programmable Guide RNA/Spacer

One sub-component is the programmable guide RNA. Due to its simplicity and efficiency, the CRISPR-Cas system can be used to perform genome-editing in cells of various organisms. The specificity of this system is dictated by base pairing between a target DNA
 5 and a custom-designed guide RNA. By engineering and adjusting the base-pairing properties of guide RNAs, one can target any sequences of interest provided that there is a PAM sequence in a target sequence.

Among the sub-components of the magRNA disclosed herein, the guide sequence provides the targeting specificity. It includes a region that is complementary and capable of
 10 hybridization to a pre-selected target site of interest. In various embodiments, this guide sequence can comprise from about 10 nucleotides to more than about 25 nucleotides. For example, the region of base pairing between the guide sequence and the corresponding target site sequence can be about 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25, or more than 25 nucleotides in length. In an exemplary embodiment, the guide sequence is about 17-
 15 20 nucleotides in length, such as 20 nucleotides.

In some embodiments, one requirement for selecting a suitable target nucleic acid is that it has a 3' PAM site/sequence. In some examples, each target sequence and its corresponding PAM site/sequence are referred herein as a Cas-targeted site. The Type II CRISPR system, one of the most well characterized systems, needs only Cas9 protein and a
 20 guide RNA complementary to a target sequence to affect target cleavage. The type II CRISPR system of *S. pyogenes* uses target sites having N12-20NGG, where NGG represents the PAM site from *S. pyogenes*, and N12-20 represents the 12-20 nucleotides directly 5' to the PAM site. Additional PAM site sequences from other species of bacteria include NGGNG, NNNNGATT, NNAGAA, NNAGAAW, and NAAAAC. See, e.g., US 20140273233, WO 2013176772,
 25 Cong *et al.*, (2012), Science 339 (6121): 819–823, Jinek *et al.*, (2012), Science 337 (6096): 816–821, Mali *et al.*, (2013), Science 339 (6121): 823–826, Gasiunas *et al.*, (2012), Proc Natl Acad Sci U S A. 109 (39): E2579–E2586, Cho *et al.*, (2013) Nature Biotechnology 31, 230–232, Hou *et al.*, Proc Natl Acad Sci U S A. 2013 Sep 24;110(39):15644-9, Mojica *et al.*, Microbiology. 2009 Mar;155(Pt 3):733-40, and www.addgene.org/CRISPR/. The contents of
 30 these documents are incorporated herein by reference in their entireties.

The target nucleic acid strand can be either of the two strands on a genomic DNA in a host cell. Examples of such genomic dsDNA include, but are not necessarily limited to, a host cell chromosome, mitochondrial DNA and a stably maintained extrachromosomal DNA.

However, it is to be understood that the present method can be practiced on other dsDNA present in a host cell, such as non-stable plasmid DNA, viral DNA, and phagemid DNA, as long as there is Cas-targeted site regardless of the nature of the host cell dsDNA. The present method can be practiced on RNAs too.

5 *RNA scaffold capable of binding to the RNA-guided nickase*

Besides the above-described guide sequence, the magRNA may include additional active or non-active sub-components. In one example, the magRNA may have an RNA motif (e.g. CRISPR motif) with tracrRNA activity. For example, the magRNA can be a hybrid RNA molecule where the above-described programmable guide RNA is fused to a tracrRNA to
10 mimic the natural crRNA:tracrRNA duplex.

In some embodiments, an magRNA: comprise both a crRNA and a tracrRNA, with matching sequencing following tracrRNA. The matching tag can be in the crRNA or TracrRNA.

Various tracrRNA sequences are known in the art and examples include the following
15 tracrRNAs and active portions thereof. In one example, an active portion of a tracrRNA retains the ability to form a complex with a Cas protein, such as Cas9 or dCas9. See, e.g., WO2014144592. Methods for generating crRNA-tracrRNA hybrid RNAs are known in the art. See e.g., WO2014099750, US 20140179006, and US 20140273226. The contents of these documents are incorporated herein by reference in their entireties. In some embodiments, the
20 tracrRNA activity and the guide sequence are two separate RNA molecules, which together form the guide RNA and related scaffold. In this case, the molecule with the tracrRNA activity should be able to interact with (usually by base pairing) the molecule having the guide sequence.

Matching tag or anchoring tag

25 One unique feature of ME is the magRNA, which contains a matching tag (also called anchoring tag) complementary to a segment in the modular RNA template. In some embodiments, the tag in magRNA is at the 3' end of magRNA. In other embodiments, the tag is at the 5' end of magRNA. In some embodiments, the tag and spacer are at the opposite ends of the RNA scaffold; in some embodiments, the anchoring tag and spacer are at the same end
30 of the RNA scaffold.

In some embodiments, the length of the tag is 6nt, 9nt, 12nt, 15nt, 18nt, 21 nt., and 24 nt. In some embodiments, the length of the tag is of any size between 6nt and 24nt.

In some embodiments, the tag is complementary to the 5' end of the RNA template. In some embodiments, the tag matches an internal sequence in the RNA template. In some embodiments, the 5' end complementary sequence in the RNA template is 23nt, 29nt, 35nt, 41nt, 50nt, 100nt upstream (5' to) the reverse transcription initiation site. In some

5 embodiments, the complementary sequence in the RNA template sequence is of a distance between 20nt to 10kb upstream the reverse transcription initiation site.

In some embodiments, when a dual-module ME with two magRNAs is used, the two tag sequences in two different magRNA are complementary to two different segments in the RNA template.

10 In some embodiments when a single-module ME is used, a match tag matching a sequence 29nt upstream the reverse transcription initiation site exhibits high efficacy.

In some embodiment, two magRNA with two different tags are used to recruit two separate RNA templates.

The selection of magRNA tag and RNA template complimentary region pair is

15 governed by Watson-Crick base pairing. GC content, length, and melting points are taken consideration to design a strong anchoring magRNA tag.

Examples of RNA scaffolds and magRNA are listed below:

sgRNA scaffold encoding sequence (T will be transcribed to U in RNA) (spCas9)

5' GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTG
20 AAAAAGTGGCACCGAGTCGGTGC-3' (SEQ ID NO: 33)

A "TT" is sometimes added to the 3' end to stabilize magRNA.

Split gRNA/magRNA sequences (containing a crRNA and transRNA)

25 crRNA (spCas9): 5' - NNNNNNNNNNNNNNNNNNNNNN GUUUUAGAGCUAUGCU-3'
(N20 is guide/spacer sequence) (SEQ ID NO: 34)

tracrRNA-mag (spCas9)

5' - AGCAUAGCA AGUUA AAAUA AGGCU AGUCC GUUAU CAACU UGAAA
30 AAGUG GCACC GAGUC GGUGC-MMMMMMMMMMMMM-3' (SEQ ID NO: 35) (M12
is the match sequence). UU is sometime added before the first M.

sgRNA scaffold for saCas9

5' GTTTTAGTACTCTGGAAACAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGT
35 CAACTTGTTGGCGAGA-3' (SEQ ID NO: 36)

magRNA match sequences for EGFP-200L Guide

3' magRNA match sequences for EGFP-200L Guide
(gctgaagcactgcacgccgC/T) (SEQ ID NO: 37)

Name	3' match sequence

12M-Ex23	5'-gtgggtcacgagg-3' (SEQ ID NO: 38)
12M-Ex29	5'-acgaggggtgggc-3' (SEQ ID NO: 39)
12M-Ex35	5'-gtggggccagggc-3' (SEQ ID NO: 40)
12M-Ex41	5'-cagggcacgggc-3' (SEQ ID NO: 41)
12M-Ex57	5'-tgccggtggtgc-3' (SEQ ID NO: 42)
6M-Ex29	5'-acgagg-3'
9M-Ex29	5'-acgaggggtg-3'
15M-Ex29	5'-acgaggggtgggccag-3' (SEQ ID NO: 43)
18M-Ex29	5'-acgaggggtgggccagggc-3' (SEQ ID NO: 44)
21M-Ex29	5'-acgaggggtgggccagggcacg-3' (SEQ ID NO: 45)
24M-Ex29	5'-acgaggggtgggccagggcacgggc-3' (SEQ ID NO: 46)
5' match magRNA match sequences for EGFP-200L Guide (gctgaagcactgcacgccgC/T)	
Name	5' match sequence
5'mag	5'-acgaggggtgggc-3' (SEQ ID NO: 47)
5'mag-15nt_linker	5'acgaggggtgggcAAATTAACAACTAA3' (SEQ ID NO: 48)
second magRNA for Dual-ME Dual-Mag	
Name	5' Guide 3' match sequence
13MEx1 13 (EGF P-92U)	5'ggccacaagtgcagcgtgtc-3' (SEQ ID NO: 49)
13MEx1 11 (EGF P119U)	5'ggcgagggcgatgccaccta-3' (SEQ ID NO: 51)
	5'-gccctcgccggac-3' (SEQ ID NO: 50)
	5'-tcgccctcgccgg-3' (SEQ ID NO: 52)

Recruiting RNA Motif

In some embodiments, the magRNA may include a recruiting RNA motif(s), which links or recruits a reverse transcriptase or an effector protein to a target site.

- 5 One way to recruit a reverse transcriptase or an effector protein to a target sequence is through a direct fusion to an RNA-guided nickase, such as dCas9. The direct fusion of effector to the proteins required for sequence recognition (such as dCas9) has achieved success in sequence specific transcriptional activation or suppression, but the protein-protein fusion design may render spatial hindrance, which is not ideal for proteins (*e.g.*, enzymes) that need
- 10 to form a multimeric complex for their activities. In that case, it is advantageous to recruit a

reverse transcriptase or an effector protein to the target site in a *split* manner as shown in FIGs. 6, 7, 8, 10B, 11B, 12B, 14, and 16.

More specifically, this *split* system takes advantage of various RNA motif/RNA binding protein binding pairs. To this end, the magRNA can be designed such that an RNA motif (*e.g.*, MS2 operator motif), which specifically binds to an RNA binding protein (*e.g.*, MS2 coat protein, MCP), is linked or incorporated into it. The recruiting RNA motif can be fused to any suitable place in the magRNA. For example, it could replace the loops within the RNA scaffold, specifically the tetraloop and/or stem loop.

As a result, this RNA scaffold component of the magRNA disclosed herein is a designed RNA molecule, which contains not only a gRNA motif for specific DNA/RNA sequence recognition (*e.g.*, a CRISPR RNA motif for dCas9 binding), but also the recruiting RNA motif for effector recruitment. In this way, recruited RT or effector protein fusions can be recruited to the target site through their ability to bind to the recruiting RNA motif. Due to the flexibility of RNA scaffold mediated recruitment, a functional monomer, as well as dimer, tetramer, or oligomer could be relatively easy to form near the target DNA or RNA sequence. These pairs of RNA recruiting motif/binding protein could be derived from naturally occurring sources (*e.g.*, RNA phages, or yeast telomerase) or could be artificially designed (*e.g.*, RNA aptamers and their corresponding binding protein ligands). A non-exhaustive list of examples of recruiting RNA motif/RNA binding protein pairs that could be used in the system described herein is summarized in Table 2.

Table 2. Examples of recruiting RNA motifs that can be used in this disclosure, as well as their pairing RNA binding proteins/protein domains.

RNA motif	Pairing interacting protein*	Organism
Telomerase Ku binding motif	Ku	Yeast
Telomerase Sm7 binding motif	Sm7	Yeast
MS2 phage operator stem-loop	MS2 Coat Protein (MCP)	Phage
PP7 phage operator stem-loop	PP7 coat protein (PCP)	Phage
SfMu phage Com stem-loop	Com RNA binding protein	Phage
Non-natural RNA aptamer	Corresponding aptamer ligand	Artificially designed

The sequences for the above binding pairs are listed below.

1. Telomerase Ku binding motif / Ku heterodimer

a. Ku binding hairpin

5' –

UUCUUGUCGUACUUAUAGAU CGCUACGUUAUUUCAAUUUUGAAAAUCUGAGUCCUGGGAGUG CGGA–3' (SEQ ID NO: 53)

b. Ku heterodimer

MSGWESYYKTEGDEEAEQEEENLEASGDYKYSGRDSLIFLVDASKAMFESQSEDELTPFDM
 SIQCIQSVYISKIISDRDLLAVVFGTEKDKNSVNFKNIIYVLQELDNPGAKRILELDQFKG
 QQGQKRFQDMMGHGS DYSLSEVLWVCANLFSQVQFKMSHKRIMLFTNEDNPHGND SAKASRA
 RTKAGDLRDTGIFLDLMLHKKPGGFDISLFYRDIISIAEDEDLRVHFEESSKLEDLLRKVRA
 5 KETRKRALSRLKLNKDIVISVGIYNLVQKALKPPPIKLYRETNEPVKTKTRTFNTSTGGL
 LLPSDTKRSQIYGSRQIILEKEETEELKRFDDPGLMLMGFKPLVLLKKHHYLRPSLFVYPPEE
 SLVIGSSTLFSALLIKCLEKEVAALCRYTPRRNIPPYFVALVPQEEELDDQKIQVTPPGFQL
 VFLPFADDKRKMPFTEKIMATPEQVGKMKAI VEKLRFTYRSDSFENPVLQQHFRNLEALALD
 LMEPEQAVDLTLPKVEAMNKRGLSLVDEFKELVYPPDYNPEGKVTKRKHDNEGSGSKRPKVE
 10 YSEEELKTHISKGTGLGKFTVPMLEACRAYGLKSGLKKQELLEALT KHFQD (SEQ ID
 NO: 54)

MVRSGNKA AVVLCMDVGFTMSNSIPGIESPF EQAKKVI TMFVQRQVFAENKDEIALVLFGTD
 GTDNPLSGGDQYQNI TVHRHMLLPDFDLLEDIESKI QPGSQQADFLDALIVSMDVIQHETIG
 15 KKF EKRHIEIF TDLSSRFSKSQLDIIHSLKKCDISERHSIHWPCRLTIGSNLSIRIAAYKS
 ILQERVKKTWTVD AKTLKKEDIQKETVYCLNDDDETEVLKEDI IQGFRYGS DIVPF SKVDE
 EQMKYKSE GKCF SVLGFC KSSQVQRRFFMG NQVLKVFAARDDEAAVALSSLIHALDDLD MV
 AIVRYAYDKRANPQVGVA FPHIKHNYECLVYVQLPFMEDLRQYMFSS LKNSK KYAPTEAQLN
 AVDALIDSM SLAKKDEKTD TLEDLFPTTKIPNPRFQRLFQCLLHRALHPREPLPPIQQHIWN
 20 MLNPPAEVTTKSQIPLSKI KTLFPLIEAKKKDQVTAQE I FQDNHEDGPTAK (SEQ ID
 NO: 55)

2. Telomerase Sm7 binding motif / Sm7 homoheptamer

a. Sm consensus site (single stranded)
 25 5'-AAUUUUUGGA-3' (SEQ ID NO: 56)

b. Monomeric Sm - like protein (archaea)
 GSVIDVSSQRVNVQRPLDALGNSLNSPVI IKLKGDREFRGVLKSF DLHMNLVLNDAAEELEDG
 EVTRRLGTVLIRGDNIVYISP (SEQ ID NO: 57)

3. MS2 phage operator stem loop / MS2 coat protein

a. MS2 phage operator stem loop
 5' - GCGCACAUGAGGAUCACCAUGUGC -3' (SEQ ID NO: 58)

b. MS2 coat protein
 MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIK
 35 VEVPGAWRSYLN MELTIPIFATNSDC ELIVKAMQGLLKDGNPIPSAIAANS GIY (SEQ
 ID NO: 59)

4. PP7 phage operator stem loop / PP7 coat protein

a. PP7 phage operator stem loop
 40 5'-AUAAGGAGUUUAUAUGGAAACCCUUA -3' (SEQ ID NO: 60)

b. PP7 coat protein (PCP)
 MSKTIVLSVGEATRTLTEIQSTADRQIFEEKVGPLVGRRLRTASLRQNGAKTAYRVNLKLDQ
 ADVVDCSTSVCGELPKVRYTQVWSDVTIVANSTEASRKSLYDLTKSLVATSQVEDLVVNLV
 (SEQ ID NO: 61)

5. SfMu Com stem loop / SfMu Com binding protein

a. SfMu Com stem loop
 5'-CUGAAUGCCUGCGAGCAUC-3' (SEQ ID NO: 62)

b. SfMu Com binding protein
 50 MKSIRCKNCNKL LFKADSFDHIEIRCPRCKRHIIMLNACEHPTEKHCGKREKITHSDETVRY
 (SEQ ID NO: 63)

6. sgRNA scaffold with MS2 inserted at Stem-Loop (spCas9)

5' GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGCCAACA
TGAGGATCACCCATGTCTGCAGGGCCAAGTGGCACCGAGTCGGTGC-3' (SEQ ID NO: 64)

7. sgRNA scaffold with MS2 inserted at Stem-Loop for saCas9

5' GTTTTAGTACTCTGGAAACAGAATCTACTAAAACAAGGCAAAATGCCGTGTTT
ATCTCGTCAAGGCCAACATGAGGATCACCCATGTCTGCAGGGCCTTGGCGAGA-3'
(SEQ ID NO: 65)

RNA template molecule

One key feature of the ME system is the separation of the template RNA from gRNA, and freedom in template RNA design. The modular RNA template only needs to contain a short track polyribonucleotide sequence at its 3' end that is complementary to nicking DNA strand with a free 3' end, so the nicking DNA can serve as a primer for the new DNA strand synthesis. The length of the primer binding site (PBS) is customary between 9 nt to 18 nt. Primer binding site longer than 18 is also acceptable.

As the tag sequence in magRNA depends on the sequence of the RNA template, not the other way around, there is no need to consider including any exogenous sequences in the template for recruitment. The design is solely for introducing a desired new sequence change into the target location of the genome. In some embodiments, point mutations, either transitions or transversions, are included in the RNA template for introducing point mutation. In some embodiments, a deletion is included in the RNA template for introducing deletion. In some embodiments, an insertion is included for introducing insertion. In some embodiments 1nt, 4nt, 29nt, 52nt changes, including point mutations, deletions, and insertions, are introduced. Other lengths of nucleotide alteration between 1 nt to 52 nt can also be made. In some embodiments, mutations, insertions, and deletions of the size between 53 nt-100nt are introduced. In some embodiments, insertion and deletion of the size between 101nt to 50kb are programmed by the RNA template.

In some embodiments, the distance between the sequence change to the nicking site is 1nt, 4nt, 20nt, 35nt, 50nt, 100nt. The other distance between 1nt to 100nt can also be selected. In some embodiments the distance is between 100nt and 1000nt.

In some embodiments, 32nt, 41nt, 47nt, 72nt, 131nt, 135nt RNA templates are effective for ME. In some embodiments, the modular RNA template is of a size between 15nt and 200 nt. In some embodiments, the RNA template is of a size between 201 nt and 2 kb. In some embodiments, the RNA template is of a size between 2 kb to 10 kb.

Customarily, the RNA template contains a stretch or stretches of sequences that are identical or have homology to sequences at vicinity of the target site to facility cellular repair and increase editing efficiency.

5 The gene editing complex or system described herein can transcribe an RNA sequence template into host target DNA sites by target-primed reverse transcription. By writing a DNA sequence via reverse transcription of the RNA sequence template directly into the host genome, the gene editing complex or system can insert an object sequence into a target genome without the need for exogenous DNA sequences to be introduced into the host cell, as well as eliminate an exogenous DNA insertion. Therefore, the gene editing complex or system provides a
10 platform for the use of customized RNA sequence templates containing object sequences, *e.g.*, sequences comprising heterologous gene coding and/or function information.

In some embodiments, an RNA template may comprise an open reading frame or the reverse complement thereof. In some embodiments, an RNA template may comprise a coding sequence for a bacterial or viral antigen epitope. In some embodiments, an RNA template may
15 comprise a sequence encoding antigen recognition variable region of antibody. In some embodiments, an RNA template may comprise a sequence encoding T cell receptor (TCR) antigen recognition variable region. In some embodiments, and RNA template may comprise a sequence for a signal peptide for excretion or a nuclear localization signal for nuclear transport, or a peptide for intracellular trafficking. In some embodiments the RNA template
20 may be converted into double stranded DNA (*e.g.*, through reverse transcription) before the open reading frame can be transcribed and translated. In some embodiments, the RNA may comprise homology to the DNA target site.

In certain embodiments, a RNA template can be identified, designed, engineered and constructed to contain sequences altering or specifying host genome function, for example by
25 introducing a heterologous coding region into a genome; affecting or causing exon structure/alternative splicing; causing disruption of an endogenous gene; causing transcriptional activation of an endogenous gene; causing epigenetic regulation of an endogenous DNA; causing up- or down-regulation of operably linked genes, etc. In certain embodiments, an RNA template can be engineered to contain sequences coding for exons
30 and/or transgenes, provide for binding sites to transcription factor activators, repressors, enhancers, etc., and combinations of thereof. In other embodiments, the coding sequence can be further customized with splice acceptor sites, poly-A tails.

The RNA template may have some homology to the target DNA. In some embodiments the RNA template has 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 or more bases of exact homology to the target DNA at the 3' end of the RNA template. In some embodiments the RNA template has 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, 160, 175, 180, or 200 or more bases of at least 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 98%, 99% or 100% homology to the target DNA, *e.g.*, at the 5' end of the RNA template.

The RNA template component described herein typically is able to bind to the magRNA. In some embodiments the RNA template has a region that is capable of binding to the magRNA. The region can be in any suitable position within the RNA template.

The RNA template typically comprises an object sequence for insertion into a target DNA. The object sequence may be coding or non-coding. In some embodiments a system or method described herein comprises a single RNA template. In some embodiments a system or method described herein comprises a plurality of RNA templates.

In some embodiments, the template contains mutations intended to change the PAM or proto-spacer sequences required by the magRNA or gRNA. As a result, while the CRISPR complex can edit the original DNA sequence, the edited DNA sequence is not subject to recognition by the CRISPR complex, avoiding futile repetitive editing. In some embodiments, the template contains mutations to generate new PAMs or proto-spacer, allowing CRISPR complex to only target the edited sequences but not the unedited sequences, when desired.

In some embodiments, the object sequence may contain an open reading frame. In some embodiments the RNA template has a Kozak sequence. In some embodiments the RNA template has an internal ribosome entry site. In some embodiments the RNA template has a self-cleaving peptide such as a T2A or P2A site. In some embodiments the RNA template has a start codon. In some embodiments the RNA template has a splice acceptor site. In some embodiments the RNA template has a splice donor site. In some embodiments the RNA template has a stop codon. In some embodiments the RNA template has a microRNA binding site downstream of the stop codon. In some embodiments the RNA template has a polyA tail downstream of the stop codon of an open reading frame. In some embodiments the RNA template comprises one or more exons. In some embodiments the RNA template comprises one or more introns. In some embodiments the RNA template comprises a eukaryotic

transcriptional terminator. In some embodiments the RNA template comprises an enhanced translation element or a translation enhancing element.

In some embodiments the RNA template comprises coding sequences for functional small non-protein coding RNAs. Examples of these functional non-coding RNAs (ncRNA) include transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), as well as small RNAs such as microRNAs, siRNAs, piRNAs, snoRNAs, snRNAs, exRNAs, scaRNAs. In some embodiments the non-coding RNAs are engineered small regulatory RNAs that enhance or inhibit the functions of their endogenous ncRNA counterparts.

In some embodiments, the object sequence may contain a non-coding sequence. For example, the RNA template may comprise a promoter or enhancer sequence. In some embodiments the RNA template comprises a tissue specific promoter or enhancer, each of which may be unidirectional or bidirectional. In some embodiments the promoter is an RNA polymerase I promoter, RNA polymerase II promoter, or RNA polymerase III promoter. In some embodiments the promoter comprises a TATA element. In some embodiments the promoter has one or more binding sites for transcription factors.

In some embodiments, an RNA template may comprise a promoter sequence, *e.g.*, a tissue specific promoter or enhancer sequence. In some embodiments, the tissue-specific promoter or enhancer is used to increase the target-cell specificity of a gene. For instance, the promoter or enhancer can be chosen on the basis that it is active in a target cell type but not active in (or active at a lower level in) a non-target cell type. Thus, even if the promoter or enhancer integrated into the genome of a non-target cell, it would not drive expression (or only drive low level expression) of an integrated gene. A system having a tissue-specific promoter or enhancer sequence in the RNA template may also be used in combination with a microRNA binding site, *e.g.*, in the RNA template, *e.g.*, as described herein. In some embodiments the RNA template may comprise a microRNA sequence, a siRNA sequence, a guide RNA sequence, a piwi RNA sequence. In some embodiments a tissue specific silencer or repressor sequence is used to silence or repress the expression of a target gene in the target cells and tissue.

In some embodiments the RNA template may comprise a site that coordinates epigenetic modification. In some embodiments the RNA template comprises an element that inhibits, *e.g.*, prevents, epigenetic silencing. In some embodiments the RNA template comprises a chromatin insulator. For example, the RNA template can comprise a CTCF site or a site targeted for DNA methylation.

In order to promote higher level or more stable gene expression, the RNA template may include features that prevent or inhibit gene silencing. In some embodiments, these features prevent or inhibit DNA methylation. In some embodiments the RNA template comprises a sequence mediating transcriptional repression. In some embodiments, these features promote DNA demethylation. In some embodiments, these features prevent or inhibit histone deacetylation. In some embodiments, these features prevent or inhibit histone methylation. In some embodiments, these features promote histone acetylation. In some embodiments, these features promote histone demethylation. In some embodiments, multiple features may be incorporated into the RNA template to promote one or more of these modifications. CpG dinucleotides are subject to methylation by host methyl transferases. In some embodiments, the RNA template is depleted of CpG dinucleotides, *e.g.*, does not comprise CpG nucleotides or comprises a reduced number of CpG dinucleotides compared to a corresponding unaltered sequence. In some embodiments, the promoter driving transgene expression from integrated DNA is depleted of CpG dinucleotides.

In some embodiments the RNA template comprises a gene expression unit composed of at least one regulatory region operably linked to an effector sequence. The effector sequence may be a sequence that is transcribed into RNA (*e.g.*, a coding sequence or a non-coding sequence such as a sequence encoding a micro RNA).

In some embodiments the object sequence of the RNA template can be, *e.g.*, 50-50,000 base pairs (*e.g.*, between 50-40,000 bp, between 500-30,000 bp between 500-20,000 bp, between 100-15,000 bp, between 500-10,000 bp, between 50-10,000 bp, between 50-5,000 bp. In some embodiments, the heterologous object sequence is less than 1,000, 1,300, 1500, 2,000, 3,000, 4,000, 5,000, or 7,500 nucleotides in length.

Modifications

The RNA molecules (*e.g.*, magRNA or RNA template) described herein may include one or more modifications. Such modifications may include inclusion of at least one non-naturally occurring nucleotide, or a modified nucleotide, or analogs thereof. Modified nucleotides may be modified at the ribose, phosphate, and/or base moiety. Modified nucleotides may include 2'-O-methyl analogs, 2'-deoxy analogs, or 2'-fluoro analogs. The nucleic acid backbone may be modified, for example, a phosphorothioate backbone may be used. The use of locked nucleic acids (LNA) or bridged nucleic acids (BNA) may also be possible. Further examples of modified bases include, but are not limited to, 2-aminopurine, 5-bromo-uridine, pseudouridine, inosine, 7-methylguanosine. These modifications may apply

to any component of the system described herein. In a preferred embodiment these modifications can be made to the RNA components, *e.g.*, the guide RNA sequence.

In some embodiments, the RNA molecule described above or a subsection thereof can comprise one or more modifications, *e.g.*, a base modification, a backbone modification, etc,
 5 to provide the nucleic acid with a new or enhanced feature (*e.g.*, improved stability).

Modified Backbones and Modified Inter-nucleoside Linkages

Examples of suitable nucleic acids containing modifications include nucleic acids containing modified backbones or non-natural internucleoside linkages. Nucleic acids (having modified backbones include those that retain a phosphorus atom in the backbone and those that
 10 do not have a phosphorus atom in the backbone.

Suitable modified oligonucleotide backbones containing a phosphorus atom therein include, for example, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates, 5'-alkylene phosphonates and chiral phosphonates, phosphinates,
 15 phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, phosphorodiamidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, selenophosphates and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein one or more internucleotide linkages is a 3' to 3', 5' to 5' or 2' to 2' linkage. Suitable oligonucleotides having
 20 inverted polarity comprise a single 3' to 3' linkage at the 3'-most internucleotide linkage *i.e.* a single inverted nucleoside residue which may be a basic (the nucleobase is missing or has a hydroxyl group in place thereof). Various salts (such as, for example, potassium or sodium), mixed salts and free acid forms are also included.

In some embodiments, a subject nucleic acid comprises one or more phosphorothioate
 25 and/or heteroatom internucleoside linkages, in particular —CH₂—NH—O—CH₂—, —CH₂—N(CH₃)—O—CH₂—(known as a methylene (methylimino) or MMI backbone), —CH₂—O—N(CH₃)—CH₂—, —CH₂—N(CH₃)—N(CH₃)—CH₂— and —O—N(CH₃)—CH₂—CH₂— (wherein the native phosphodiester internucleotide linkage is represented as —O—P(=O)(OH)—O—CH₂—). MMI type internucleoside linkages are disclosed in the above
 30 referenced U.S. Pat. No. 5,489,677. Suitable amide internucleoside linkages are disclosed in t U.S. Pat. No. 5,602,240.

Also suitable are nucleic acids having morpholino backbone structures as described in, *e.g.*, U.S. Pat. No. 5,034,506. For example, in some embodiments, a subject nucleic acid

comprises a 6-membered morpholino ring in place of a ribose ring. In some of these embodiments, a phosphorodiamidate or other non-phosphodiester internucleoside linkage replaces a phosphodiester linkage.

Suitable modified polynucleotide backbones that do not include a phosphorus atom therein have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; riboacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH₂ component parts.

Mimetics

A subject nucleic acid (*e.g.*, an RNA template, a magRNA, and others) disclosed herein can be a nucleic acid mimetic. The term "mimetic" as it is applied to polynucleotides is intended to include polynucleotides wherein only the furanose ring or both the furanose ring and the internucleotide linkage are replaced with non-furanose groups, replacement of only the furanose ring is also referred to in the art as being a sugar surrogate. The heterocyclic base moiety or a modified heterocyclic base moiety is maintained for hybridization with an appropriate target nucleic acid. One such nucleic acid, a polynucleotide mimetic that has been shown to have excellent hybridization properties, is referred to as a peptide nucleic acid (PNA). In PNA, the sugar-backbone of a polynucleotide is replaced with an amide containing backbone, in particular an aminoethylglycine backbone. The nucleotides are retained and are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone.

One polynucleotide mimetic that has been reported to have excellent hybridization properties is a peptide nucleic acid (PNA). The backbone in PNA compounds is two or more linked aminoethylglycine units which gives PNA an amide containing backbone. The heterocyclic base moieties are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone. Representative U.S. patents that describe the preparation of PNA compounds include, but are not limited to: U.S. Pat. Nos. 5,539,082; 5,714,331; and 5,719,262.

In some embodiments, a DNA template is used in place of the RNA template, and a DNA polymerase is used in place of the reverse transcriptase.

Another class of polynucleotide mimetic that has been studied is based on linked morpholino units (morpholino nucleic acid) having heterocyclic bases attached to the morpholino ring. A number of linking groups have been reported that link the morpholino monomeric units in a morpholino nucleic acid. One class of linking groups has been selected to give a non-ionic oligomeric compound. The non-ionic morpholino-based oligomeric compounds are less likely to have undesired interactions with cellular proteins. Morpholino-based polynucleotides are non-ionic mimics of oligonucleotides which are less likely to form undesired interactions with cellular proteins (Dwayne A. Braasch and David R. Corey, Biochemistry, 2002, 41(14), 4503-4510). Morpholino-based polynucleotides are disclosed in U.S. Pat. No. 5,034,506. A variety of compounds within the morpholino class of polynucleotides have been prepared, having a variety of different linking groups joining the monomeric subunits.

A further class of polynucleotide mimetic is referred to as cyclohexenyl nucleic acids (CeNA). The furanose ring normally present in a DNA/RNA molecule is replaced with a cyclohexenyl ring. CeNA DMT protected phosphoramidite monomers have been prepared and used for oligomeric compound synthesis following classical phosphoramidite chemistry. Fully modified CeNA oligomeric compounds and oligonucleotides having specific positions modified with CeNA have been prepared and studied (see Wang *et al.*, J. Am. Chem. Soc., 2000, 122, 8595-8602). In general the incorporation of CeNA monomers into a DNA chain increases its stability of a DNA/RNA hybrid. CeNA oligoadenylates formed complexes with RNA and DNA complements with similar stability to the native complexes. The study of incorporating CeNA structures into natural nucleic acid structures was shown by NMR and circular dichroism to proceed with easy conformational adaptation.

A further modification includes Locked Nucleic Acids (LNAs) in which the 2'-hydroxyl group is linked to the 4' carbon atom of the sugar ring thereby forming a 2'-C,4'-C-oxymethylene linkage thereby forming a bicyclic sugar moiety. The linkage can be a methylene ($\text{—CH}_2\text{—}$), group bridging the 2' oxygen atom and the 4' carbon atom wherein n is 1 or 2 (Singh *et al.*, Chem. Commun., 1998, 4, 455-456). LNA and LNA analogs display very high duplex thermal stabilities with complementary DNA and RNA ($T_m = +3$ to $+10^\circ\text{C}$.), stability towards 3'-exonucleolytic degradation and good solubility properties. Potent and nontoxic antisense oligonucleotides containing LNAs have been described (Wahlestedt *et al.*, Proc. Natl. Acad. Sci. U.S.A., 2000, 97, 5633-5638).

The synthesis and preparation of the LNA monomers adenine, cytosine, guanine, 5-methyl-cytosine, thymine and uracil, along with their oligomerization, and nucleic acid recognition properties have been described (Koshkin *et al.*, Tetrahedron, 1998, 54, 3607-3630). LNAs and preparation thereof are also described in WO 98/39352 and WO 99/14226.

Modified Sugar Moieties

A subject nucleic acid can also include one or more substituted sugar moieties. Suitable polynucleotides comprise a sugar substituent group selected from: OH; F; O-, S-, or N-alkyl; O-, S-, or N-alkenyl; O-, S- or N-alkynyl; or O-alkyl-Co-alkyl, wherein the alkyl, alkenyl and alkynyl may be substituted or unsubstituted C₁ to C₁₀ alkyl or C₂ to C₁₀ alkenyl and alkynyl.

Particularly suitable are O((CH₂)_nO)_mCH₃, O(CH₂)_nOCH₃, O(CH₂)_nNH₂, O(CH₂)_nCH₃, O(CH₂)_nONH₂, and O(CH₂)_nON((CH₂)_nCH₃)₂, where n and m are from 1 to about 10. Other suitable polynucleotides comprise a sugar substituent group selected from: C₁ to C₁₀ lower alkyl, substituted lower alkyl, alkenyl, alkynyl, alkaryl, aralkyl, O-alkaryl or O-aralkyl, SH, SCH₃, OCN, Cl, Br, CN, CF₃, OCF₃, SOCH₃, SO₂CH₃, ONO₂, NO₂, N₃, NH₂, heterocycloalkyl, heterocycloalkaryl, aminoalkylamino, polyalkylamino, substituted silyl, an RNA cleaving group, a reporter group, an intercalator, a group for improving the pharmacokinetic properties of an oligonucleotide, or a group for improving the pharmacodynamic properties of an oligonucleotide, and other substituents having similar properties. A suitable modification includes 2'-methoxyethoxy (2'-O—CH₂CH₂OCH₃, also known as 2'-O-(2-methoxyethyl) or 2'-MOE) (Martin *et al.*, Helv. Chim. Acta, 1995, 78, 486-504) *i.e.*, an alkoxyalkoxy group. A further suitable modification includes 2'-dimethylaminoethoxyethoxy, *i.e.*, a O(CH₂)₂ON(CH₃)₂ group, also known as 2'-DMAOE, as described in examples hereinbelow, and 2'-dimethylaminoethoxyethoxy (also known in the art as 2'-O-dimethyl-amino-ethoxy-ethyl or 2'-DMAEOE), *i.e.*, 2'-O—CH₂—O—CH₂—N(CH₃)₂.

Other suitable sugar substituent groups include methoxy (—O—CH₃), aminopropoxy (—OCH₂CH₂NH₂), allyl (—CH₂—CH=CH₂), —O-allyl CH₂—CH=CH₂) and fluoro (F). 2'-sugar substituent groups may be in the arabino (up) position or ribo (down) position. A suitable 2'-arabino modification is 2'-F. Similar modifications may also be made at other positions on the oligomeric compound, particularly the 3' position of the sugar on the 3' terminal nucleoside or in 2'-5' linked oligonucleotides and the 5' position of 5' terminal nucleotide. Oligomeric compounds may also have sugar mimetics such as cyclobutyl moieties in place of the pentofuranosyl sugar.

Base Modifications and Substitutions

A subject nucleic acid may also include nucleobase (often referred to in the art simply as “base”) modifications or substitutions. As used herein, “unmodified” or “natural” nucleobases include the purine bases adenine (A) and guanine (G), and the pyrimidine bases thymine (T), cytosine (C) and uracil (U). Modified nucleobases include other synthetic and natural nucleobases such as 5-methylcytosine (5-me-C), 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-thiouracil, 2-thiothymine and 2-thiocytosine, 5-halouracil and cytosine, 5-propynyl ($\text{—C}\equiv\text{C—CH}_3$) uracil and cytosine and other alkynyl derivatives of pyrimidine bases, 6-azo uracil, cytosine and thymine, 5-uracil (pseudouracil), 4-thiouracil, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, 8-hydroxyl and other 8-substituted adenines and guanines, 5-halo particularly 5-bromo, 5-trifluoromethyl and other 5-substituted uracils and cytosines, 7-methylguanine and 7-methyladenine, 2-F-adenine, 2-amino-adenine, 8-azaguanine and 8-azaadenine, 7-deazaguanine and 7-deazaadenine and 3-deazaguanine and 3-deazaadenine. Further modified nucleobases include tricyclic pyrimidines such as phenoxazine cytidine (1H-pyrimido(5,4-b)(1,4)benzoxazin-2(3H)-one), phenothiazine cytidine (1H-pyrimido(5,4-b)(1,4)benzothiazin-2(3H)-one), G-clamps such as a substituted phenoxazine cytidine (*e.g.* 9-(2-aminoethoxy)-H-pyrimido(5,4-b)(1,4)benzoxazin-2(3H)-one), carbazole cytidine (2H-pyrimido(4,5-b)indol-2-one), pyridoindole cytidine (H-pyrido(3',2':4,5)pyrrolo(2,3-d)pyrimidin-2-one).

Heterocyclic base moieties may also include those in which the purine or pyrimidine base is replaced with other heterocycles, for example 7-deaza-adenine, 7-deazaguanosine, 2-aminopyridine and 2-pyridone. Further nucleobases include those disclosed in U.S. Pat. No. 3,687,808, those disclosed in The Concise Encyclopedia Of Polymer Science And Engineering, pages 858-859, Kroschwitz, J. I., ed. John Wiley & Sons, 1990, those disclosed by Englisch *et al.*, Angewandte Chemie, International Edition, 1991, 30, 613, and those disclosed by Sanghvi, Y. S., Chapter 15, Antisense Research and Applications, pages 289-302, Crooke, S. T. and Lebleu, B., ed., CRC Press, 1993. Certain of these nucleobases are useful for increasing the binding affinity of an oligomeric compound. These include 5-substituted pyrimidines, 6-azapyrimidines and N-2, N-6 and O-6 substituted purines, including 2-aminopropyladenine, 5-propynyluracil and 5-propynylcytosine. 5-methylcytosine substitutions have been shown to increase nucleic acid duplex stability by 0.6-1.2 °C. (Sanghvi *et al.*, eds., Antisense Research

and Applications, CRC Press, Boca Raton, 1993, pp. 276-278) and are suitable base substitutions, *e.g.*, when combined with 2'-O-methoxyethyl sugar modifications.

Expression System

To use the complex, system, or platform described above, it may be desirable to express one or more of the protein and RNA components from nucleic acids that encode them. This can be performed in a variety of ways. For example, the nucleic acids encoding the RNAs or proteins can be cloned into one or more intermediate vectors for introducing into prokaryotic or eukaryotic cells for replication and/or transcription. Intermediate vectors are typically prokaryotic vectors, *e.g.*, plasmids, or shuttle vectors, or insect vectors, for storage or manipulation of the nucleic acid encoding the RNAs or protein for production of the RNA or protein. The nucleic acids can also be cloned into one or more expression vectors, for administration to a plant cell, animal cell, preferably a mammalian cell or a human cell, fungal cell, bacterial cell, or protozoan cell. Accordingly, the present disclosure provides nucleic acids that encode any of the RNAs or proteins mentioned above. Preferably, the nucleic acids are isolated and/or purified.

The present disclosure also provides recombinant constructs or vectors having sequences encoding one or more of the RNAs or proteins described above. Examples of the constructs include a vector, such as a plasmid or viral vector, into which a nucleic acid sequence of the disclosure has been inserted, in a forward or reverse orientation. In a preferred embodiment, the construct further includes regulatory sequences, including a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are also described in *e.g.*, Sambrook *et al.* (2001, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press).

A vector refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. The vector can be capable of autonomous replication or integration into a host DNA. Examples of the vector include a plasmid, cosmid, or viral vector. The vector of this disclosure includes a nucleic acid in a form suitable for expression of the nucleic acid in a host cell. Preferably, the vector includes one or more regulatory sequences operatively linked to the nucleic acid sequence to be expressed. A "regulatory sequence" includes promoters, enhancers, and other expression control elements (*e.g.*, polyadenylation signals). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence, as well as inducible regulatory sequences. The design of the expression vector can

depend on such factors as the choice of the host cell to be transformed, transfected, or transduced, the level of expression of RNAs or proteins desired, and the like.

Examples of expression vectors include chromosomal, non-chromosomal and synthetic DNA sequences, bacterial plasmids, phage DNA, baculovirus, yeast plasmids, vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies. However, any other vector may be used provided it is replicable and viable in the host. The appropriate nucleic acid sequence may be inserted into the vector by a variety of procedures. In general, a nucleic acid sequence encoding one of the RNAs or proteins described above can be inserted into an appropriate restriction endonuclease site(s) by procedures known in the art. Such procedures and related sub-cloning procedures are within the scope of those skilled in the art.

The vector may include appropriate sequences for amplifying expression. In addition, the expression vector preferably contains one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell cultures, or such as tetracycline or ampicillin resistance in *E. coli*.

The vectors for expressing the RNAs can include RNA Pol III promoters to drive expression of the RNAs, *e.g.*, the HI, U6 or 7SK promoters. The vectors for expressing the RNAs can include Pol I promoters to drive expression of the template RNAs, *e.g.*, influenza viral promoters. These human promoters allow for expression of RNAs in mammalian cells following vector introduction. Alternatively, a T7 promoter may be used, *e.g.*, for *in vitro* transcription, and the RNA can be transcribed *in vitro* and purified.

The vector containing the appropriate nucleic acid sequences as described above, as well as an appropriate promoter or control sequence, can be employed to transform, transfect, or infect an appropriate host to permit the host to express the RNAs or proteins described above. Examples of suitable expression hosts include bacterial cells (*e.g.*, *E. coli*, *Streptomyces*, *Salmonella typhimurium*), fungal cells (yeast), insect cells (*e.g.*, *Drosophila* and *Spodoptera frugiperda* (Sf9)), animal cells (*e.g.*, CHO, COS, and HEK 293), adenoviruses, and plant cells. The selection of an appropriate host is within the scope of those skilled in the art. In some embodiments, the present disclosure provides methods for producing the above mentioned RNAs or proteins by transforming, transfecting, or infecting a host cell with an expression vector having a nucleotide sequence that encodes one of the RNAs, or polypeptides, or proteins.

The host cells are then cultured under a suitable condition, which allows for the expression of the RNAs or proteins.

Any of the procedures known in the art for introducing foreign nucleotide sequences and proteins into host cells may be used. Examples include the use of calcium phosphate transfection, polybrene, protoplast fusion, electroporation, nucleofection, liposomes, microinjection, naked DNA, plasmid vectors, viral vectors, both episomal and integrative, and any of the other well-known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell.

In some embodiments the components of the match editing systems can be delivered in the form of RNA (*e.g.* mRNA encoding the proteins, magRNA, and template RNA), DNA (DNA expression vectors), or protein (*e.g.*, purified CRISPR-RT fusion protein), or any combination of the above forms (*e.g.*, ribonucleoprotein complex).

Any of the procedures known in the art for deliver the foreign nucleotide sequences and proteins systemically into host may be used. Examples include the use of viral vectors, both episomal and integrative, such as Adeno-associated virus (AAV), adenoviral vectors (AD), lentiviral vectors, retroviral vectors, integration deficient lentiviral vectors, Herpes simplex virus vectors (HSV), stomatitis virus vectors (VSV), modified vaccinia virus Ankara vectors (MVA), arenavirus viral vectors, Sendai virus vectors, parvovirus vectors. Examples include the use of non-viral vectors, such as liposome, lipid particles, lipid nanoparticles, genome-free versions virus-like particles (VLPs).

Methods

Another aspect of the present disclosure encompasses a method for modifying a target DNA sequence (*e.g.*, a chromosomal sequence) or target RNA sequence in a cell, embryo, human or non-human animals. The sequence modifications can be base changes, insertions, deletions, or a combination of these modifications. In one embodiment, the embryo is a non-human animal embryo. The method comprises introducing into the cell or embryo the above-described (A) an RNA-guided nickase; (B) a reverse transcriptase; (C) an RNA template molecule and (D) a magRNA molecule. The magRNA guides the other components to a target polynucleotide at a target site and the reverse transcriptase write sequence into the target site based on sequence of the RNA template molecule as described herein. The RNA template contains the desired nucleotide sequence modification information.

The target polynucleotide herein is the sequence where an RNA-guided nickase creates a single strand DNA break, or nicking. The target polynucleotide nicking site has no sequence

limitation except that the sequence is immediately followed (downstream or 3') by a PAM sequence if a CRISPR type nickase is used. Examples of PAM include, but are not limited to, NG, NGG, NGGNG, and NNAGAAW (wherein N is defined as any nucleotide and W is defined as either A or T). Other examples of PAM sequences are given above, and the skilled person will be able to identify further PAM sequences for use with a given nickase (*e.g.*, a CRISPR protein). In some examples the RNA-guided nickases may not require a PAM motif (*e.g.*, the artificially engineered PAM-less CRISPR proteins). The target nucleotide modification can be in the coding region of a gene, in an intron of a gene, in a transcriptional regulatory region of a gene, in a control region between genes, etc. The gene can be a protein-coding gene or an RNA coding gene.

The desired nucleotide sequence changes and the location of the changes are dictated by the designed RNA template. The RNA template is an individual RNA molecule that is not covalently linked to the guide RNA. The RNA template can be designed freely depending on the desired sequence changes to be made, without the concerns of interference with guide RNA structure. There is no distance requirement between the desired DNA modification and the target nicking site. The distance anywhere within a 200- nucleotide range exhibits high editing efficacy. There is no content requirement for the RNA template design except for comprising a primer binding sequence at the 3' end complementary to the nicked DNA strand. There is no requirement for including any exogenous appendicular sequence in the RNA template for recruitment. This freedom in designing the RNA template mechanistically eliminated the fundamental source of "scars" that are often incorporated into the genome editing site when other reverse transcriptase, recombinase, integrase, and transposase editing tools are used.

AFTER the RNA template is designed for a desired DNA modification at a desired target site, an RNA tag is then designed that is complementary to a sequence segment within the RNA template. The tag is added to the gRNA, usually at the 3' end, to recruit and anchor the RNA template to the target nicking site. The size of the RNA tag can be between 6 nucleotides long and 24 nucleotide long. The position within the RNA template where the tag RNA is complementary to can be at 5' end or within the RNA template.

In another example, this disclosure provides a system and a related method that recruit one or more additional, different functional effectors to the same target sequence. The effectors work together synergistically to facilitate the genetic conversion. Examples can include proteins that facilitate the editing process or increase editing efficiency, such as chromatin modification enzymes, inhibitors of mismatch repair enzymes. The examples of the effectors

can be a human RNase Inhibitor protein (RNH1), a dominant negative protein of MMR pathway protein including MLH1 protein, a 5' DNA nuclease Fen1 protein, or the like that could enhance the editing efficiency.

Example of a human RNase inhibitor-MCP fusion protein sequence (NLS-RNH1-

5 Linker25-MCP):

PKKKRKVMSLDIQCEELSDARWAEELLPLLQQCQVVRLLDDCGLTEARCKDISSALRVNPALAE
 LNLRSNELGDVGVHCVLQGLQTPSCKIQKLSLQNCCLTGAGCGVLSSTLRTLPTLQELHLSD
 NLLGDAGLQLLCEGLLDPQCRLEKLQLEYCSLSAASCEPLASVLRAPDFKELTVSNNDINE
 AGVRVLCQGLKDSPCQLEALKLESCGVTSDNCRDLGIVASKASLRELALGSNKLGDVGMAE
 10 LCPGLLHPSSRLRTLWIWECGITAKGCGDLCRVLRAKESLKELSLAGNELGDEGARLLCETL
 LEPGCQLESLWVKSCSFTAACCSHFSSVLAQNRFLLELQISNNRLEDAGVRELQGLGQPGS
 VLRVLWLADCDVSDSSCSLAATLLANHSLRELDLSNNCLGDAGILQLVESVRQPGCLLEQL
 VLYDIYWSEEMEDRLQALEKDKPSLRVISELKTPLGDTTHTSPPCPAPELLGGPMASNFTQF
 VLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW
 15 RSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNIPIPSAIAANSGLY (SEQ ID NO: 66)

The match editing system thus provides an extremely versatile method to edit polynucleotide in a cell. The cell can be a single-cell organism, or cells isolated from a human or non-human organism, or cells derived for or engineered from a human or non-human
 20 organism, or cells in a human or non-human organism.

The target polynucleotide to be edited can be any polynucleotide endogenous or exogenous to the cell. For example, the target polynucleotide can be a DNA molecule 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). The
 25 DNA molecule can be endogenous to the cell, or of infected virus or microbe.

The protein components of this system of this disclosure can be introduced into the cell or embryo as an isolated protein. Alternatively, the components can be introduced via nucleic acids encoding such components, such DNA or RNA (*e.g.*, in vitro transcribed RNA). In one embodiment, each protein can comprise at least one cell-penetrating domain, which facilitates
 30 cellular uptake of the protein. In other embodiments, mRNA molecules or DNA molecules encoding the protein or proteins can be introduced into the cell or embryo. In general, a DNA sequence encoding the protein is operably linked to a promoter sequence that will function in the cell or embryo of interest. The DNA sequence can be linear, or the DNA sequence can be part of a vector. In still other embodiments, the protein can be introduced into the cell or
 35 embryo as an RNA-protein complex comprising the protein, the RNA template, and the magRNA described above.

In alternate embodiments, DNA encoding the protein(s) can further comprise a sequence or sequences encoding components of the RNA component (*e.g.*, the magRNA and/or the RNA template). In general, the DNA sequence encoding the protein and the RNA is operably linked to appropriate promoter control sequences that allow the expression of the protein and the RNA, respectively, in the cell or embryo. The DNA sequence encoding the protein and the RNA can further comprise additional expression control, regulatory, and/or processing sequence(s). The DNA sequence encoding the protein and the RNA can be linear or can be part of a vector.

In embodiments in which the RNA is introduced into the cell via a DNA molecule encoding the RNA, the RNA coding sequence can be operably linked to promoter control sequence for expression of the RNA in the eukaryotic cell. For example, the RNA coding sequence can be operably linked to a promoter sequence that is recognized by RNA polymerase III (Pol III). The RNA coding sequence can be operably linked to a promoter sequence that is recognized by RNA polymerase I (Pol I). Examples of suitable Pol III promoters include, but are not limited to, mammalian U6 or H1 promoters. In exemplary embodiments, the RNA coding sequence is linked to a mouse or human U6 promoter. In other exemplary embodiments, the RNA coding sequence is linked to a mouse or human H1 promoter. In other embodiments, the RNA coding sequence is linked to a viral Pol I promoter such as an influenza Pol I promoter.

The DNA molecule encoding the protein and/or RNA described herein can be linear or circular. In some embodiments, the DNA sequence can be part of a vector, such as a multicistronic vector. Suitable vectors include plasmid vectors, phagemids, cosmids, artificial/minichromosomes, transposons, and viral vectors. In an exemplary embodiment, the DNA encoding the protein and/or RNA is present in a plasmid vector. Non-limiting examples of suitable plasmid vectors include pUC, pBR322, pET, pBluescript, and variants thereof. The vector can comprise additional expression control sequences (*e.g.*, enhancer sequences, Kozak sequences, polyadenylation sequences, transcriptional termination sequences, etc.), selectable marker sequences (*e.g.*, antibiotic resistance genes), origins of replication, and the like.

The protein components of this system of this disclosure (or nucleic acid(s) encoding them) and the RNA components (or DNAs encoding them) can be introduced into a cell or embryo by a variety of means. In one embodiment, the embryo is a non-human animal embryo. Typically, the embryo is a fertilized one-cell stage embryo of the species of interest. In some embodiments, the cell or embryo is transfected. Suitable transfection methods include calcium

phosphate-mediated transfection, nucleofection (or electroporation), cationic polymer transfection (*e.g.*, DEAE-dextran or polyethylenimine), viral transduction, virosome transfection, virion transfection, liposome transfection, cationic liposome transfection, immunoliposome transfection, nonliposomal lipid transfection, dendrimer transfection, heat shock transfection, magnetofection, lipofection, gene gun delivery, impalefection, sonoporation, optical transfection, gold nanoparticle-mediated transfection, and proprietary agent-enhanced uptake of nucleic acids. Transfection methods are well known in the art (see, *e.g.*, "Current Protocols in Molecular Biology" Ausubel *et al.*, John Wiley & Sons, New York, 2003 or "Molecular Cloning: A Laboratory Manual" Sambrook & Russell, Cold Spring Harbor Press, Cold Spring Harbor, N.Y., 3rd edition, 2001). In other embodiments, the molecules are introduced into the cell or embryo by microinjection. For example, the molecules can be injected into the pronuclei of one-cell embryos.

The protein components of this system disclosed herein (or nucleic acid(s) encoding them) and the RNA components (or DNAs encoding them) can be introduced into a cell or embryo simultaneously or sequentially. The ratio of the protein (or its encoding nucleic acid) to the RNA (or DNAs encoding the RNA), generally will be approximately stoichiometric such that they can form an RNA-protein complex. Similarly, the ratio of two different proteins (or encoding nucleic acids) will be approximately stoichiometric. In one embodiment, the protein components and the RNA components (or the DNA sequences encoding them) are delivered together within the same nucleic acid or vector.

The method further comprises maintaining the cell or embryo under appropriate conditions such that the guide RNA guides the effector protein to the targeted site in the target sequence, and the effector domain modifies the target sequence.

In general, the cell can be maintained under conditions appropriate for cell growth and/or maintenance. Suitable cell culture conditions are well known in the art and are described, for example, in "Current Protocols in Molecular Biology" Ausubel *et al.*, John Wiley & Sons, New York, 2003 or "Molecular Cloning: A Laboratory Manual" Sambrook & Russell, Cold Spring Harbor Press, Cold Spring Harbor, N.Y., 3rd edition, 2001), Santiago *et al.* (2008) PNAS 105:5809-5814; Moehle *et al.* (2007) PNAS 104:3055-3060; Urnov *et al.* (2005) Nature 435:646-651; and Lombardo *et al.* (2007) Nat. Biotechnology 25:1298-1306. Those of skill in the art appreciate that methods for culturing cells are known in the art and can and will vary depending on the cell type. Routine optimization may be used, in all cases, to determine the best techniques for a particular cell type.

An embryo can be cultured *in vitro* (e.g., in cell culture). Typically, the embryo is cultured at an appropriate temperature and in appropriate media with the necessary O₂/CO₂ ratio to allow the expression of the proteins and RNA scaffold, if necessary. Suitable non-limiting examples of media include M2, M16, KSOM, BMOC, and HTF media. A skilled artisan will appreciate that culture conditions can and will vary depending on the species of embryo. Routine optimization may be used, in all cases, to determine the best culture conditions for a particular species of embryo. In some cases, a cell line may be derived from an *in vitro*-cultured embryo (e.g., an embryonic stem cell line).

Alternatively, an embryo may be cultured *in vivo* by transferring the embryo into a uterus of a female host. Generally speaking, the female host is from the same or similar species as the embryo. Preferably, the female host is pseudo-pregnant. Methods of preparing pseudo-pregnant female hosts are known in the art. Additionally, methods of transferring an embryo into a female host are known. Culturing an embryo *in vivo* permits the embryo to develop and can result in a live birth of an animal derived from the embryo. Such an animal would comprise the modified chromosomal sequence in every cell of the body.

A variety of eukaryotic cells are suitable for use in the method. For example, the cell can be a human cell, a non-human mammalian cell, a non-mammalian vertebrate cell, an invertebrate cell, an insect cell, a plant cell, a yeast cell, or a single cell eukaryotic organism. A variety of embryos are suitable for use in the method. For example, the embryo can be a 1-cell, 2-cell, or 4-cell human or non-human mammalian embryo. Exemplary mammalian embryos, including one cell embryos, include without limit mouse, rat, hamster, rodent, rabbit, feline, canine, ovine, porcine, bovine, equine, and primate embryos. In still other embodiments, the cell can be a stem cell. Suitable stem cells include without limit embryonic stem cells, ES-like stem cells, fetal stem cells, adult stem cells, pluripotent stem cells, induced pluripotent stem cells, multipotent stem cells, oligopotent stem cells, unipotent stem cells and others. In exemplary embodiments, the cell is a mammalian cell, or the embryo is a mammalian embryo.

A variety of mammalian cells are suitable for match editing includes cells isolated from a human or non-human subject, or cells derived or engineered from a human or non-human subject, including NK cells, pluripotent stem cells (PSCs), adult stem cells (ASCs), fibroblasts, chondrocytes, keratinocytes, hepatocytes, pancreatic islet cells, and immune cells, including T cells, dendritic cells (DCs), natural killer (NK) cells, and macrophages.

Cells are suitable for Match Editing also include cells in a human and non-human subject. In some examples, the Match Editing components can be delivered systemically using viral or non-viral vectors. Any of the procedures known in the art for deliver the foreign nucleotide sequences and proteins systemically into host may be used. Examples include the use of viral vectors, both episomal and integrative, such as Adeno-associated virus (AAV), adenoviral vectors (AD), lentiviral vectors, retroviral vectors, integration deficient lentiviral vectors, Herpes simplex virus vectors (HSV), stomatitis virus vectors (VSV), modified vaccinia virus Ankara vectors (MVA), arenavirus viral vectors, Sendai virus vectors, parvovirus vectors. Examples include the use of non-viral vectors, such as liposome, lipid particles, lipid nanoparticles, genome-free versions virus-like particles (VLPs).

Utilities and Applications

The systems and methods disclosed herein have a wide variety of utilities including modifying and editing (*e.g.*, inactivating and activating) a target polynucleotide in a multitude of cell types. As such the systems and methods have a broad spectrum of applications in, *e.g.*, research and therapy. For example, the systems and methods can be used for high throughput screening where multiple systems with different guide RNAs target multiple different loci to obtain and screen for multiple different phenotypic outcomes (*e.g.*, better proliferation or lethal screens in cell lines). In another example, the systems and methods can be used in mutagenesis (similar to CRISPR tiling) or genes to create novel proteins.

In some embodiments, the above-described complex, system, composition, or method of genome editing can be used for generating point mutation (both transversion and transition), insertion, deletion in cells derived from an organism of plant and animal organisms, and humans.

In some embodiments, the above-described complex, system, composition, or method of genome editing can be used for generating point mutation (both transversion and transition), insertion, deletion in cells within an organism of plant and animal organisms, and humans.

The above-described complex, system, composition, or method of genome editing can be used in cell engineering, cell therapy, organism engineering, gene therapy, agricultural improvement, veterinary medicine, and cellular and animal models for research.

When used for making mutation, insertion, and deletion, the match editing technology disclosed herein enables precision genetic manipulation of DNA by copying a desired sequence from an RNA molecule and inserting it into a target position in the genome. This technology can effectively introduce genetic changes including a single nucleotide change, both transition

or transversion, insertion, and deletion, or a combination of these genetic alterations. Importantly, the new features of Match Editing technology render independent modular design of RNA template and the magRNA. Consequently, this design avoids the potential interference between the gRNA and template secondary structures, eliminates the sources of genetic “scars” of the template, and makes synergistic multiplexing possible and convenient.

In some embodiments, the match editing disclosed herein, through installing single base changes, insertions, or deletions, can correct causal mutations of genetic diseases. In some embodiments, the match editing through installing single base changes, insertions, or deletions, can lead to introduction of stop codons, generating missense frameshift, or eliminating splicing sites. As a result, the expression of the target gene can be effectively silenced. In a broad sense, the technology allows rewire of cellular regulatory circuitry.

In one example, the match editing disclosed herein can change or install transcriptional regulatory sequences such as transcription factor binding sites, eliminating or introducing specific transcriptional activation or inhibition. In another example, it can change protein modification sites such as phosphorylation sites or acetylation sites, changing cellular signal transduction. In another example, it can change or install amino acid interface involved in protein-protein or protein-nucleic acid interaction, eliminating or installing specific protein interaction with other macromolecules, re-wire cellular signaling pathways.

In some embodiments, ME can insert a bacterial or viral antigen epitope into cellular protein. In some embodiments, ME can insert an antigen recognition variable region of antibody to a protein. In some embodiments, ME can replace T cell receptor (TCR) antigen recognition variable region. In some embodiments, ME can insert a signal peptide for excretion or a nuclear localization signal for nuclear transport, or a peptide for intracellular trafficking.

In some embodiments, the match editing disclosed herein, through inserting or replacing a short stretch of peptide coding sequence within a target gene, can specifically “label” an intracellular protein or membrane protein by creating a tag.

In some embodiments, the match editing disclosed herein, through inserting or replacing a peptide sequence within a cell, can change cell surface protein properties, generating new cell-cell interacting patterns. In one example, T-cell receptor (TCR) protein variable sequences can be altered to re-direct the T cells for targeting different antigen presenting cells. In another example, the B-Cell receptor (BCR) protein variable sequences can be changed to re-direct the B cell targets.

In some embodiments, the match editing disclosed herein can be used to change a sequence of a secretory protein are changed to make cells that can secrete altered hormones, cytokines, and growth factors. In some examples, the secretory protein can be an antibody, where the antibody gene variable regions can be changed for antibody production switch.

5 In some embodiments, the match editing disclosed herein can be used to engineer neuronal cells so that the cells can be differentially tagged and their interactions and wiring can be re-designed and monitored.

The match editing disclosed herein can be broadly used for a number of important areas.

10 It can be used for research, therapeutic development, agriculture development, industrial organismal engineering.

For therapeutic development, Match Editing can be used for ex vivo therapeutic development through engineering therapeutic cells, either autologous cells or allogeneic cells.

15 In one example, it can be used for engineering autologous hematopoietic cells from disease-inflicted patients. For example, hematopoietic cells from patients with sickle cell anemia or beta thalassemia can be engineered ex vivo either to correct the underlying disease-causing mutations or to reactivate the fetal hemoglobin expression by changing the regulatory sequence of the gene. The engineered autologous cells will then be infused back to the patients.

20 The similar strategy of introducing point mutations or deletion at the transcriptional regulatory elements of a gene can be generally applied for treating other diseases either through eliminating transcriptional repression, or introducing transcriptional activation, or introducing transcriptional repression, all through changing the transcription factor binding sites.

25 In one example, Match editing can introduce stop codon and frameshift insertion/deletion to inactivate genes for cell therapy engineering. For example, for generating allogeneic CAR-T cells, genes involved in graft vs host diseases (GvHD) and host vs graft reaction (HvGR) as well as genes negatively affecting CAR-T functionality can be inactivated.

30 In one example, GvHD and HvGR genes and genes negatively impacting therapeutic cell functions can be inactivated in other types of therapeutic cells, including NK cells, pluripotent stem cells (PSCs), adult stem cells (ASCs), fibroblasts, chondrocytes, keratinocytes, hepatocytes, pancreatic islet cells, and immune cells, including T cells, dendritic cells (DCs), natural killer (NK) cells, and macrophages derived from a human or non-human subject, for generating allogeneic therapeutic cells and more potent therapeutic cells.

In one example, the Match Editing technology can replace the MHC-antigen complex recognition sequence of the T cell receptor with sequence recognizing MHC- cancer antigen

complex sequence for generating TCR-T cells for cancer cell therapies. Moreover, allogeneic cells can also be generated with above mentioned strategies for CAR-T cells. Furthermore, the TCR various region replacement by match editing can also be achieved for in vivo therapy.

In some examples, the Match Editing technology can install an active cis-transcriptional enhancer element to the FOXP3 gene in T cells to generate regulatory T cells (Treg). In some examples, the TCR in Treg cells is further engineered by Match Editing method for tissue and cell specific targeting. In some examples the Treg cells are further engineered by Match Editing on the HLA loci and MHC genes for generating allogeneic Treg. In some examples, the Treg cell engineering steps are carried out simultaneously with multiplexing Match Editing or in combination with other gene editing approaches.

For therapeutic development, match editing components or expression vectors for genes encoding the components can be delivered with appropriate delivery vehicles including viral and non-viral delivery systems for in vivo gene therapy.

In one example, the in vivo therapy is to deliver the match editing system in vivo for correcting the delta 508 deletion of CFTR gene for treating cystic fibrosis.

In one example, the in vivo therapy is to deliver the match editing system in vivo for exon skipping of the mutated dystrophin gene to partially restore dystrophin function for the treatment of Duchenne Muscular Dystrophy.

In one example, the in vivo therapy is to deliver the match editing system in vivo to reduce the expression of amyloid beta ($A\beta$), tau, or α -synuclein for treating Alzheimer's disease.

The similar principle of ex vivo cell engineering for cell therapy and in vivo gene therapy can apply to the vast majority human genetic diseases because the versatile of the gene editing strategy of match editing can cover point mutations, insertions, deletions, and the combination of these mutations. Moreover, the mutation sites do not need to be immediately adjacent to the PAM.

The similar principle of ex vivo cell engineering for cell therapy and in vivo gene therapy can apply to scavenge or inactivate expression of disease-causing proteins or RNA for treating cancers, autoimmune diseases, and neurodegenerative diseases.

In one example, the autologous therapeutic cells can be engineered; in another example, allogeneic therapeutic cells can be engineered by inactivating the GvHD and HvGR genes.

Match editing can be used for engineering of iPSC from human or non-human origin, for the treatment of human and veterinary diseases.

Match editing can be used for engineering germline cells of non-human mammals, non-human animals, plants, agriculture plants. Live organisms can be produced from these engineered germline cells.

Match editing can be used for industrial and non-industrial microbe genome engineering.

Many devastating human diseases have one common cause: genetic alteration or mutation. The disease-causing mutations in patients either are acquired through inheritance from their parents or are caused by environmental factors. These diseases include, but are not limited to, the following categories. First, some genetic disorders are caused by germline mutations. One example is cystic fibrosis, which is caused by mutations at the CFTR gene inherited from parents. A second suppressor mutation in the mutant CFTR can partially restore the function of CFTR protein in somatic tissues. Other example genetic diseases caused by a point genetic mutation that can be corrected include Gaucher's disease, alpha trypsin deficiency disease, sickle cell anemia, to name a few. Second, some diseases, such as chronic viral infectious diseases, are caused by exogenous environmental factors and resulting genetic alterations. One example is AIDS, which is caused by insertion of the human HIV viral genome into the genome of infected T-cells. Third, some neurodegenerative diseases involve genetic alterations. One example is Huntington's diseases, which is caused by expansion of CAG trinucleotide in the huntingtin gene of affected patients. Other examples include lysosomal storage diseases, Epidermolysis Bullosa, and retinal degeneration. Finally, cancers are caused by various somatic mutations accumulated in cancer cells. Therefore, correcting the disease-causing genetic mutations, or functionally correcting the sequence, provides an appealing therapeutic opportunity to treat these diseases.

Somatic genetic editing is an appealing therapeutic strategy for many human diseases. To achieve successful therapeutic genetic editing, three critical factors are considered essential: (i) how to achieve sequence specific recognition ("sequence recognition module"); (ii) how to correct the underlying mutations ("correction module"); and (iii) how to link the "correction module" to "sequence recognition module" together to achieve sequence specific correction. There is a number of ways of achieving each individual task. However, none of the currently existing platforms or technologies could achieve optimal and practical somatic genetic editing. More specifically, current gene specific editing technologies are mostly based on nucleases induced DNA DSB and consequent DSB induced homologous recombination, the activity of

which is low or absent in most somatic cells. Thus, those technologies are of limited use for therapeutic corrections of pathological genetic mutations in somatic tissues in most diseases.

In contrast, the system and method disclosed in this disclosure allow DNA-sequence directed editing of a gene that does not rely on nuclease activity. The system and method do not generate DSB, or do not rely on the DSB-mediated homologous recombination. Moreover, this design of the system is modular, which allows extremely flexible and convenient way of targeting any desirable DNA or RNA sequences. In essence, this approach enables one to guide a DNA or RNA editing enzyme to virtually any DNA or RNA sequence in somatic cells, including stem cells. Through precise editing of the target DNA or RNA sequence, the enzyme can correct the mutated genes in genetic disorders, inactivate the viral genome in the infected cells, generate a stop codon for inactivation and eliminate the expression of the disease-causing protein in diseases including neurodegenerative diseases, silence the oncogenic protein in cancers, mutate a splicing consensus cite to eliminate a disease causing exon, or mutate a regulatory sequence to restore a therapeutic expression/inactivation of a gene. Accordingly, the system and method disclosed in this disclosure can be used in correcting underlying genetic alterations in diseases including the above-mentioned genetic disorders, chronic infectious diseases, neurodegenerative diseases, and cancer. Importantly, the system and method disclosed in this disclosure can be used to engineer cells, for both generating research tools or for generating cell-based therapies.

Genetic Diseases

It is estimated that over six thousand genetic diseases are caused by known genetic mutations. Correcting the underlying disease causing mutations in the pathological tissues/organs can provide alleviation or cure to the diseases. For example, cystic fibrosis affects 1 out of every 3,000 people in the US. It is caused by inheritance of a mutated CFTR gene and 70% of the patients have the same mutation, deletion of a tri-nucleotide leading to a deletion of phenylalanine at position 508 (called Δ Phe 508). Δ Phe 508 leads to the mislocation and degradation of CFTR. The system and method disclosed in this disclosure can be used to convert a Val 509 residue (GTT) to Phe 509 (TTT) in affected tissues (lung), thereby functionally correct the Δ Phe 508 mutation. In addition, a second suppressor mutation (such as R553Q or R553M or V510D) in the mutant Δ Phe 508 CFTR can partially restore the function of CFTR protein in somatic tissues.

Chronic Infectious Diseases

The system and method disclosed in this disclosure can also be used to specifically inactivate any gene in a viral genome that is incorporated into human cells/tissues. For example, the system and method disclosed in this disclosure allow one to create a stop codon for early termination of translation of the essential viral genes, and thereby remediate or cure the chronic debilitating infectious diseases. For example, current AIDS therapies can reduce viral load, but cannot totally eliminate dormant HIV from positive T cells. The system and method disclosed herein can be used to permanently inactivate expression of essential HIV genes in the integrated HIV genome in human T-cells by introducing one or multiple stop codons. Another example is hepatitis B virus (HBV). The system and method disclosed here can be used to specifically inactivate essential HBV genes, which are incorporated into human genome, and silence the HBV life cycle.

Neurodegenerative Diseases

Some neurodegenerative diseases are caused by gain-of-function mutations. For example, SOD1G93A leads to development of amyotrophic lateral sclerosis (ALS). The system and method disclosed in this disclosure can be used to either correct the mutation or eliminate the mutant protein expression by introducing a stop codon or by changing a splice site. For example, an alternative splicing form of Tau protein that includes exon 10 plays a causal role in Alzheimer's disease. Changing a C-G base pair at the consensus exon 10 splice site would abolish the alternative splicing version of Tau.

Cancers

Many genes (including tumor suppressor genes, oncogenes, and DNA repair genes) contribute to the development of cancer. Mutations in these genes often lead to various cancers. Using the system and method disclosed in this disclosure, one can specifically target and correct these mutations. As a result, causative oncogenic proteins can be functionally repressed or their expression can be eliminated by introducing a point mutation at either the catalytic sites or splicing sites.

Somatic gene knockout

In some embodiments, protein expression of a gene in somatic cells in human and non-human organisms can be eliminated by generating a pre-mature stop-codon. This approach can be used for therapeutic purpose or for generating research tools.

Alteration of regulatory elements

The method could be used to change sequence of regulatory elements in DNA and RNA. Consequently, it provides an approach for altering, silencing, or activating expression of a gene through altering the various mechanisms involved in gene expression. This could be used for therapeutic purpose as well as for generating research tool.

Stem Cell Genetic Modification

In some embodiments, cells that are reprogrammed to become different cell types can be genetically modified using the system and method disclosed herein. Suitable cells include, *e.g.*, stem cells (adult stem cells, embryonic stem cells, induced Pluripotent Stem cells, mesenchymal stem cells *etc.* as referenced in Stem cells: past, present, and future. Zakrzewski *et al.* Stem Cell Res Ther. 2019 Feb 26;10(1):68.) and progenitor cells (*e.g.*, cardiac progenitor cells, neural progenitor cells, *etc.*) or mature cells used for conversion into a different cell type (for example using the algorithm as referenced in Molecular Interaction Networks to Select Factors for Cell Conversion. Ouyang JF *et al.*, Methods Mol Biol. 2019;1975:333-361). Suitable cells may originate from any multicellular organism including *e.g.*, mammals (including, *e.g.* rodents, humans, horses, camels, pigs), insects, avian (including, *e.g.* chicken, duck) *etc.* Suitable host cells include *in vitro* or *ex vivo* host cells, *e.g.*, isolated host cells.

In some embodiments, the complex, system, and method disclosed herein can be used for targeted and precise genetic modification of cells or tissue *ex vivo*, correcting the underlying genetic defects. After the *ex vivo* correction, the tissues could be returned to the patients. Moreover, the technology can be broadly used in cell-based therapies for correcting genetic diseases.

The term "stem cell" refers herein to a cell that under suitable conditions is capable of differentiating into a diverse range of specialized cell types, while under other suitable conditions is capable of self-renewing and remaining in an essentially undifferentiated pluripotent state. The term "stem cell" also encompasses a pluripotent cell, multipotent cell, precursor cell and progenitor cell. Exemplary human stem cells can be obtained from hematopoietic or mesenchymal stem cells obtained from bone marrow tissue, embryonic stem cells obtained from embryonic tissue, or embryonic germ cells obtained from genital tissue of a fetus. Exemplary pluripotent stem cells can also be produced from somatic cells by reprogramming them to a pluripotent state by the expression of certain transcription factors

associated with pluripotency; these cells are called "induced pluripotent stem cells" or "iPS cells or iPS cells".

An "embryonic stem (ES) cell" is an undifferentiated pluripotent cell which is obtained from an embryo in an early stage, such as the inner cell mass at the blastocyst stage, or produced by artificial means (*e.g.* nuclear transfer) and can give rise to any differentiated cell type in an embryo or an adult, including germ cells (*e.g.* sperm and eggs).

"Induced pluripotent stem cells (iPS cells or iPS cells)" are cells generated by reprogramming a somatic cell by expressing or inducing expression of a combination of factors (herein referred to as reprogramming factors). iPS cells can be generated using fetal, postnatal, newborn, juvenile, or adult somatic cells. Factors that can be used to reprogram somatic cells to pluripotent stem cells include, for example, Oct4 (sometimes referred to as Oct3/4), Sox2, c-Myc, Klf4, Nanog, and Lin28. In some embodiments, somatic cells are reprogrammed by expressing at least two reprogramming factors, at least three reprogramming factors, at least four reprogramming factors, at least five reprogramming factors, at least six reprogramming factors, or at least seven reprogramming factors to reprogram a somatic cell to a pluripotent stem cell.

"Hematopoietic progenitor cells" or "hematopoietic precursor cells" refers to cells which are committed to a hematopoietic lineage but are capable of further hematopoietic differentiation and include hematopoietic stem cells, multipotential hematopoietic stem cells, common myeloid progenitors, megakaryocyte progenitors, erythrocyte progenitors, and lymphoid progenitors. Hematopoietic stem cells (HSCs) are multipotent stem cells that give rise to all the blood cell types including myeloid (monocytes and macrophages, granulocytes (neutrophils, basophils, eosinophils, and mast cells), erythrocytes, megakaryocytes/platelets, dendritic cells), and lymphoid lineages (T-cells, B-cells, NK-cells).

"Pluripotent stem cell" refers to a stem cell that has the potential to differentiate into all cells constituting one or more tissues or organs, or preferably, any of the three germ layers: endoderm (interior stomach lining, gastrointestinal tract, the lungs), mesoderm (muscle, bone, blood, urogenital), or ectoderm (epidermal tissues and nervous system).

As used herein, the term "somatic cell" refers to any cell other than germ cells, such as an egg, a sperm, or the like, which does not directly transfer its DNA to the next generation. Typically, somatic cells have limited or no pluripotency. Somatic cells used herein may be naturally-occurring or genetically modified.

Cell Therapies and Ex Vivo Therapies

Various embodiments of the present disclosure also provide cell lines that are produced or used in accordance with any of the other embodiments of the present disclosure for use in therapy. In one embodiment, the present disclosure is directed to methods for generating therapeutic cells such as T cells engineered to express a Chimeric Antigen Receptor (CAR-T) or T Cell Receptor (TCR-T). In one embodiment, the disclosure is directed to methods of generating therapeutic regulatory T cells (Treg). The CAR-T/TCR-T cells may be derived from primary T cells or differentiated from stem cells. Suitable stem cells include, but are not limited to, mammalian stem cells such as human stem cells, including, but not limited to, hematopoietic, neural, embryonic, induced pluripotent stem cells (iPSC), mesenchymal, mesodermal, liver, pancreatic, muscle, and retinal stem cells. Other stems cells include, but are not limited to, mammalian stem cells such as mouse stem cells, *e.g.*, mouse embryonic stem cells.

In various embodiments, the complex, system and method disclosed herein may be used to knockdown, modify or increase the expression of a single gene or multiple genes in various types of cells or cell lines, including but not limited to cells from mammals. The technology may be used for many applications, including but not limited to knock down genes to prevent graft versus host disease by making non-host cells non-immunogenic to the host or prevent host vs graft disease by making non-host cells resistant to attack by the host. These approaches are also relevant to generating allogenic (off-the-shelf) or autologous (patient specific) cell-based therapeutics. Such genes include, but are not limited to, the T Cell Receptor (TRAC), the major histocompatibility complex (MHC class I and class II) genes, including B2M, co-receptors (HLA-F, HLA-G), genes involved in the innate immune response (MICA, MICB, HCP5), inflammation (NKBBiL, LTA, TNF, LTB, LST1, NCR3, AIF1), immune receptors (LY6), heat shock proteins (HSPA1L, HSPA1A, HSPA1B), complement cascade, regulatory receptors (NOTCH4), antigen processing (TAP, HLA-DM, HLA-DO), peptide transport (RING1), increased potency or persistence (such as PD-1, CTLA-4, FOXP3 and B7), genes involved in T cell interaction with the tumour microenvironment (including but not limited to receptors of cytokines such as TGFB, Interleukin (IL)-4, IL-7, IL-2, IL-4, as well as repressors of IL-15, IL-12, IL-18, IL-2, IFNgamma), genes involved in contributing to cytokine release syndrome (including but not limited to GMCSF), genes that code for the antigen targeted by the CAR/TCR (for example endogenous CS1 where the CAR is designed against CS1) or other genes found to be beneficial to CAR-T/TCR-T or other cell based therapeutics including but

not limited to CAR-NK. CAR-B etc. See, *e.g.*, DeRenzo *et al.*, Genetic Modification Strategies to Enhance CAR T Cell Persistence for Patients With Solid Tumors. Front. Immunol., 15 February 2019.

The technology may also be used to knock down or modify genes that are involved in fratricide of immune cells, such as T cells and NK cells, or genes that alert the immune system of a patient or animal that a foreign cell, particle or molecule has entered a patient or animal, or genes encoding proteins that are current therapeutic targets used to compromise or boost an immune response, for example, CD52 and PD1, respectively.

One application is to engineer HLA alleles of bone marrow cells to increase haplotype match. The engineered cells can be used for bone marrow transplantation for treating leukemia. Another application is to engineer the negative regulatory element of fetal hemoglobin gene in hematopoietic stem cells for treating sickle cell anemia and beta-thalassemia. The negative regulatory element will be mutated and the expression of fetal hemoglobin gene is re-activated in hematopoietic stem cells, compensating the functional loss due to mutations in adult alpha or beta hemoglobin genes. A further application is to engineer iPS cells for generating allogenic therapeutic cells for various degenerative diseases including Parkinson's disease (neuronal cell loss), Type 1 diabetes (pancreatic beta cell loss). Other exemplary applications include engineering HIV infection resistant T- Cells by inactivating CCR5 gene and other genes encoding receptors required for HIV entering cells.

The technology may also be used to generate transgenic animals that can be used as disease models or for gene function studies.

As used herein, the term "immune cells" generally includes white blood cells (leukocytes) which are derived from hematopoietic stem cells (HSC) produced in the bone marrow. Examples of immune cells include, but are not limited to, lymphocytes (T cells, B cells, and natural killer (NK) cells) and myeloid-derived cells (neutrophil, eosinophil, basophil, monocyte, macrophage, dendritic cells).

The immune cells may be isolated from subjects, particularly human subjects. The immune cells can be obtained from a subject of interest, such as a subject suspected of having a particular disease or condition, a subject suspected of having a predisposition to a particular disease or condition, or a subject who is undergoing therapy for a particular disease or condition. Immune cells can be collected from any location in which they reside in the subject including, but not limited to, blood, cord blood, spleen, thymus, lymph nodes, and bone

marrow. The isolated immune cells may be used directly, or they can be stored for a period of time, such as by freezing.

The immune cells may be enriched/purified from any tissue where they reside including, but not limited to, blood (including blood collected by blood banks or cord blood banks), spleen, bone marrow, tissues removed and/or exposed during surgical procedures, and tissues obtained via biopsy procedures. Tissues/organs from which the immune cells are enriched, isolated, and/or purified may be isolated from both living and non-living subjects, wherein the non-living subjects are organ donors. In particular embodiments, the immune cells are isolated from blood, such as peripheral blood or cord blood. In some aspects, immune cells isolated from cord blood have enhanced immunomodulation capacity, such as measured by CD4- or CD8-positive T cell suppression. In specific aspects, the immune cells are isolated from pooled blood, particularly pooled cord blood, for enhanced immunomodulation capacity. The pooled blood may be from 2 or more sources, such as 3, 4, 5, 6, 7, 8, 9, 10 or more sources (*e.g.*, donor subjects).

The population of immune cells can be obtained from a subject in need of therapy or suffering from a disease associated with reduced immune cell activity. Thus, the cells can be autologous to the subject in need of therapy. Alternatively, the population of immune cells can be obtained from a donor, preferably a histocompatibility matched donor. The immune cell population can be harvested from the peripheral blood, cord blood, bone marrow, spleen, or any other organ/tissue in which immune cells reside in said subject or donor. The immune cells can be isolated from a pool of subjects and/or donors, such as from pooled cord blood.

When the population of immune cells is obtained from a donor distinct from the subject, the donor is preferably allogeneic, provided the cells obtained are subject-compatible in that they can be introduced into the subject. Allogeneic donor cells are may or may not be human-leukocyte-antigen (HLA)-compatible. To be rendered subject-compatible, allogeneic cells can be treated to reduce immunogenicity.

In some embodiment, the immune cells may be T cells (*e.g.*, regulatory T cells, CD4⁺ T cells, CD8⁺ T cells, or gamma-delta T cells), NK cells, invariant NK cells, NKT cells, stem cells (*e.g.*, mesenchymal stem cells (MSCs) or induced pluripotent stem (iPSC) cells). In some embodiments, the cells are monocytes or granulocytes, *e.g.*, myeloid cells, macrophages, neutrophils, dendritic cells, mast cells, eosinophils, and/or basophils. Also provided herein are methods of producing and engineering the immune cells as well as methods of using and administering the cells for adoptive cell therapy, in which case the cells may be autologous or

allogeneic. Thus, the immune cells may be used as immunotherapy, such as to target cancer cells.

Genetic Editing in Animals and Plants

The system and method described above can be used to generate a transgenic non-human animal or plant having one or more genetic modification of interest. In some embodiments, the transgenic non-human animal is homozygous for the genetic modification. In some embodiments, the transgenic non-human animal is heterozygous for the genetic modification. In some embodiments, the transgenic non-human animal is a vertebrate, for example, a fish (*e.g.*, zebra fish, gold fish, puffer fish, cave fish, etc.), an amphibian (frog, salamander, etc.), a bird (*e.g.*, chicken, turkey, etc.), a reptile (*e.g.*, snake, lizard, etc.), a mammal (*e.g.*, an ungulate, *e.g.*, a pig, a cow, a goat, a sheep, etc.; a lagomorph (*e.g.*, a rabbit); a rodent (*e.g.*, a rat, a mouse); a non-human primate.

The gene-editing complex, system, and method disclosed herein can be used for treating diseases in animals in a way similar to those for treating diseases in humans as described above. Alternatively, it can be used to generate knock-in animal disease models bearing specific genetic mutation for purposes of research, drug discovery, and target validation. The system and method described above can also be used for introduction of point mutations to ES cells or embryos of various organisms, for purpose of breeding and improving animal stocks and crop quality.

Methods of introducing exogenous nucleic acids into plant cells are well known in the art. Suitable methods include viral infection (such as double stranded DNA viruses), transfection, conjugation, protoplast fusion, electroporation, particle gun technology, calcium phosphate precipitation, direct microinjection, silicon carbide whiskers technology, Agrobacterium-mediated transformation and the like. The choice of method is generally dependent on the type of cell being transformed and the circumstances under which the transformation is taking place (*i.e.*, *in vitro*, *ex vivo*, or *in vivo*).

Kit

This disclosure further provides kits containing reagents for performing the above-described methods, including, *e.g.*, CRISPR/Cas guided target binding or correction reaction. To that end, one or more of the reaction components, *e.g.*, RNAs, RNA-guided nickase proteins, reverse transcriptase proteins, fusion proteins and related nucleic acids, for the methods disclosed herein can be supplied in the form of a kit for use. In one embodiment, the

kit comprises a nickase protein, a reverse transcriptase protein, or a nucleic acid encoding the protein, effector protein, one or more of the RNAs described above, a set of RNA molecules described above. In other embodiments, the kit can include one or more other reaction components. In such a kit, an appropriate amount of one or more reaction components is provided in one or more containers or held on a substrate.

Examples of additional components of the kits include, but are not limited to, one or more host cells, one or more reagents for introducing foreign nucleotide sequences into host cells, one or more reagents (*e.g.*, probes or PCR primers) for detecting expression of the RNA or protein or verifying the target nucleic acid's status, and buffers or culture media for the reactions (in 1X or concentrated forms). The kit may also include one or more of the following components: supports, terminating, modifying or digestion reagents, osmolytes, and an apparatus for detection.

The reaction components used can be provided in a variety of forms. For example, the components (*e.g.*, enzymes, RNAs, probes and/or primers) can be suspended in an aqueous solution or as a freeze-dried or lyophilized powder, pellet, or bead. In the latter case, the components, when reconstituted, form a complete mixture of components for use in an assay. The kits disclosed herein can be provided at any suitable temperature. For example, for storage of kits containing protein components or complexes thereof in a liquid, it is preferred that they are provided and maintained below 0°C, preferably at or below -20°C, or otherwise in a frozen state.

A kit or system may contain, in an amount sufficient for at least one assay, any combination of the components described herein. In some applications, one or more reaction components may be provided in pre-measured single use amounts in individual, typically disposable, tubes or equivalent containers. With such an arrangement, an RNA-guided reaction can be performed by adding a target nucleic acid, or a sample or cell containing the target nucleic acid, to the individual tubes directly. The amount of a component supplied in the kit can be any appropriate amount and may depend on the target market to which the product is directed. The container(s) in which the components are supplied can be any conventional container that is capable of holding the supplied form, for instance, microfuge tubes, microtiter plates, ampoules, bottles, or integral testing devices, such as fluidic devices, cartridges, lateral flow, or other similar devices.

The kits can also include packaging materials for holding the container or combination of containers. Typical packaging materials for such kits and systems include solid matrices

(*e.g.*, glass, plastic, paper, foil, micro-particles and the like) that hold the reaction components or detection probes in any of a variety of configurations (*e.g.*, in a vial, microtiter plate well, microarray, and the like). The kits may further include instructions recorded in a tangible form for use of the components.

5 Definition

A nucleic acid or polynucleotide refers to a DNA molecule (for example, but not limited to, a cDNA or genomic DNA) or an RNA molecule (for example, but not limited to, an mRNA), and includes DNA or RNA analogs. A DNA or RNA analog can be synthesized from nucleotide analogs. The DNA or RNA molecules may include portions that are not naturally occurring, such as modified bases, modified backbone, deoxyribonucleotides in an RNA, etc. The nucleic acid molecule can be single-stranded or double-stranded.

The term “isolated” when referring to nucleic acid molecules or polypeptides means that the nucleic acid molecule or the polypeptide is substantially free from at least one other component with which it is associated or found together in nature.

As used herein, the term “guide RNA” generally refers to an RNA molecule (or a group of RNA molecules collectively) that can bind to an RNA-guided nickase (*e.g.* a CRISPR protein) and target the RNA-guided nickase protein to a specific location within a target DNA. A guide RNA can comprise two segments: a DNA-targeting guide segment and a protein-binding segment. The DNA-targeting segment comprises a nucleotide sequence that is complementary to (or at least can hybridize to under stringent conditions) a target sequence. The protein-binding segment interacts with RNA-guided nickase (*e.g.* a CRISPR protein) such as a Cas9 or Cas9 related polypeptide. These two segments can be located in the same RNA molecule or in two or more separate RNA molecules. When the two segments are in separate RNA molecules, the molecule comprising the DNA-targeting guide segment is sometimes referred to as the CRISPR RNA (crRNA), while the molecule comprising the protein-binding segment is referred to as the trans-activating RNA (tracrRNA).

As used herein, the term “target nucleic acid” or “target” refers to a nucleic acid containing a target nucleic acid sequence. A target nucleic acid may be single-stranded or double-stranded, and often is double-stranded DNA. A “target nucleic acid sequence,” “target sequence” or “target region,” as used herein, means a specific sequence or the complement thereof that one wishes to bind to or modify using a system disclosed herein. A target sequence may be within a nucleic acid *in vitro* or *in vivo* within the genome of a cell, which may be any form of single-stranded or double-stranded nucleic acid.

A “target nucleic acid strand” refers to a strand of a target nucleic acid that is subject to base-pairing with a guide RNA as disclosed herein. That is, the strand of a target nucleic acid that hybridizes with the crRNA and guide sequence is referred to as the “target nucleic acid strand.” The other strand of the target nucleic acid, which is not complementary to the guide sequence, is referred to as the “non-complementary strand.” In the case of double-stranded target nucleic acid (*e.g.*, DNA), each strand can be a “target nucleic acid strand” to design crRNA and guide RNAs and used to practice the method of this disclosure as long as there is a suitable PAM site.

As used herein, the term “derived from” refers to a process whereby a first component (*e.g.*, a first molecule), or information from that first component, is used to isolate, derive or make a different second component (*e.g.*, a second molecule that is different from the first). For example, the mammalian codon-optimized Cas9 polynucleotides are derived from the wild type Cas9 protein amino acid sequence. Also, the variant mammalian codon-optimized Cas9 polynucleotides, including the Cas9 single mutant nickase (nCas9, such as nCas9D10A) and Cas9 double mutant null-nuclease (dCas9, such as dCas9 D10A H840A), are derived from the polynucleotide encoding the wildtype mammalian codon-optimized Cas9 protein.

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.

As used herein, the term “variant” refers to a first composition (*e.g.*, a first molecule), that is related to a second composition (*e.g.*, a second molecule, also termed a “parent” molecule). The variant molecule can be derived from, isolated from, based on or homologous to the parent molecule. For example, the mutant forms of mammalian codon-optimized Cas9 (hspCas9), including the Cas9 single mutant nickase and the Cas9 double mutant null-nuclease, are variants of the mammalian codon-optimized wild type Cas9 (hspCas9). The term variant can be used to describe either polynucleotides or polypeptides.

As applied to polynucleotides, a variant molecule can have entire nucleotide sequence identity with the original parent molecule, or alternatively, can have less than 100% nucleotide sequence identity with the parent molecule. For example, a variant of a gene nucleotide sequence can be a second nucleotide sequence that is at least 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99% or more identical in nucleotide sequence compared to the original nucleotide sequence. Polynucleotide variants also include polynucleotides comprising the entire parent polynucleotide, and further comprising additional fused nucleotide sequences. Polynucleotide

variants also includes polynucleotides that are portions or subsequences of the parent polynucleotide, for example, unique subsequences (*e.g.*, as determined by standard sequence comparison and alignment techniques) of the polynucleotides disclosed herein are also encompassed by the disclosure.

5 In another aspect, polynucleotide variants include nucleotide sequences that contain minor, trivial or inconsequential changes to the parent nucleotide sequence. For example, minor, trivial or inconsequential changes include changes to nucleotide sequence that (i) do not change the amino acid sequence of the corresponding polypeptide, (ii) occur outside the protein-coding open reading frame of a polynucleotide, (iii) result in deletions or insertions that
10 may impact the corresponding amino acid sequence, but have little or no impact on the biological activity of the polypeptide, (iv) the nucleotide changes result in the substitution of an amino acid with a chemically similar amino acid. In the case where a polynucleotide does not encode for a protein (for example, a tRNA or a crRNA or a tracrRNA), variants of that polynucleotide can include nucleotide changes that do not result in loss of function of the
15 polynucleotide. In another aspect, conservative variants of the disclosed nucleotide sequences that yield functionally identical nucleotide sequences are encompassed by the disclosure. One of skill will appreciate that many variants of the disclosed nucleotide sequences are encompassed by the disclosure.

 As applied to proteins, a variant polypeptide can have entire amino acid sequence
20 identity with the original parent polypeptide, or alternatively, can have less than 100% amino acid identity with the parent protein. For example, a variant of an amino acid sequence can be a second amino acid sequence that is at least 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or more identical in amino acid sequence compared to the original amino acid sequence.

25 Polypeptide variants include polypeptides comprising the entire parent polypeptide, and further comprising additional fused amino acid sequences. Polypeptide variants also includes polypeptides that are portions or subsequences of the parent polypeptide, for example, unique subsequences (*e.g.*, as determined by standard sequence comparison and alignment techniques) of the polypeptides disclosed herein are also encompassed by the disclosure.

30 In another aspect, polypeptide variants include polypeptides that contain minor, trivial or inconsequential changes to the parent amino acid sequence. For example, minor, trivial or inconsequential changes include amino acid changes (including substitutions, deletions and insertions) that have little or no impact on the biological activity of the polypeptide, and yield

functionally identical polypeptides, including additions of non-functional peptide sequence. In other aspects, the variant polypeptides of the disclosure change the biological activity of the parent molecule, for example, mutant variants of the Cas9 polypeptide that have modified or lost nuclease activity. One of skill will appreciate that many variants of the disclosed polypeptides are encompassed by the disclosure.

In some aspects, polynucleotide or polypeptide variants of the disclosure can include variant molecules that alter, add or delete a small percentage of the nucleotide or amino acid positions, for example, typically less than about 10%, less than about 5%, less than 4%, less than 2% or less than 1%.

As used herein, the term "conservative substitutions" in a nucleotide or amino acid sequence refers to changes in the nucleotide sequence that either (i) do not result in any corresponding change in the amino acid sequence due to the redundancy of the triplet codon code, or (ii) result in a substitution of the original parent amino acid with an amino acid having a chemically similar structure. Conservative substitution tables providing functionally similar amino acids are well known in the art, where one amino acid residue is substituted for another amino acid residue having similar chemical properties (*e.g.*, aromatic side chains or positively charged side chains), and therefore does not substantially change the functional properties of the resulting polypeptide molecule.

The following are groupings of natural amino acids that contain similar chemical properties, where a substitution within a group is a "conservative" amino acid substitution. This grouping indicated below is not rigid, as these natural amino acids can be placed in different grouping when different functional properties are considered. Amino acids having nonpolar and/or aliphatic side chains include: glycine, alanine, valine, leucine, isoleucine and proline. Amino acids having polar, uncharged side chains include: serine, threonine, cysteine, methionine, asparagine and glutamine. Amino acids having aromatic side chains include: phenylalanine, tyrosine and tryptophan. Amino acids having positively charged side chains include: lysine, arginine and histidine. Amino acids having negatively charged side chains include: aspartate and glutamate.

A "Cas9 mutant" or "Cas9 variant" refers to a protein or polypeptide derivative of the wild type Cas9 protein such as *S. pyogenes* Cas9 protein, *e.g.*, a protein having one or more point mutations, insertions, deletions, truncations, a fusion protein, or a combination thereof. It retains substantially the RNA targeting activity of the Cas9 protein. The protein or polypeptide can comprise, consist of, or consist essentially of a fragment of the wild type

protein. In general, the mutant/variant is at least 50% (*e.g.*, any number between 50% and 100%, inclusive) identical to the protein. The mutant/variant can bind to an RNA molecule and be targeted to a specific DNA sequence via the RNA molecule, and may additionally have a nuclease activity. Examples of these domains include RuvC like motifs (aa. 7-22, 759-766 and 982-989) and HNH motif (aa 837-863). See Gasiunas *et al.*, Proc Natl Acad Sci U S A. 2012 September 25; 109(39): E2579–E2586 and WO2013176772.

"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.

"Hybridization" or "hybridizing" refers to a process where completely or partially complementary nucleic acid strands come together under specified hybridization conditions to form a double-stranded structure or region in which the two constituent strands are joined by hydrogen bonds. Although hydrogen bonds typically form between adenine and thymine or uracil (A and T or U) or cytosine and guanine (C and G), other base pairs may form (*e.g.*, Adams *et al.*, The Biochemistry of the Nucleic Acids, 11th ed., 1992).

As used herein, "expression" 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, pegylation, 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.

The term "fusion polypeptide" or "fusion protein" means a protein created by joining two or more polypeptide sequences together. The fusion polypeptides encompassed in this disclosure include translation products of a chimeric gene construct that joins the nucleic acid sequences encoding a first polypeptide, *e.g.*, an RNA-binding domain, with the nucleic acid sequence encoding a second polypeptide, *e.g.*, an effector domain, to form a single open-reading frame. In other words, a "fusion polypeptide" or "fusion protein" is a recombinant protein of two or more proteins which are joined by a peptide bond or via several peptides. The fusion protein may also comprise a peptide linker between the two domains.

The term "linker" refers to any means, entity or moiety used to join two or more entities. A linker can be a covalent linker or a non-covalent linker. Examples of covalent linkers include covalent bonds or a linker moiety covalently attached to one or more of the proteins or domains to be linked. The linker can also be a non-covalent bond, *e.g.*, an organometallic bond through a metal center such as platinum atom. For covalent linkages, various functionalities can be used, such as amide groups, including carbonic acid derivatives, ethers, esters, including organic and inorganic esters, amino, urethane, urea and the like. To provide for linking, the domains can be modified by oxidation, hydroxylation, substitution, reduction etc. to provide a site for coupling. Methods for conjugation are well known by persons skilled in the art and are encompassed for use in the present disclosure. Linker moieties include, but are not limited to, chemical linker moieties, or for example a peptide linker moiety (a linker sequence). It will be

appreciated that modification which do not significantly decrease the function of the RNA-binding domain and effector domain are preferred.

As used herein, the term "conjugate" or "conjugation" or "linked" as used herein refers to the attachment of two or more entities to form one entity. A conjugate encompasses both peptide-small molecule conjugates as well as peptide-protein/peptide conjugates.

The terms "subject" 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.

In some embodiments, a subject may be an invertebrate animal, for example, an insect or a nematode; while in others, a subject may be a plant or a fungus.

As used herein, "treatment" or "treating," or "palliating" or "ameliorating" are used interchangeably. These terms refer to an approach for obtaining beneficial or desired results including but not limited to a therapeutic benefit and/or a prophylactic benefit. By therapeutic benefit is meant any therapeutically relevant improvement in or effect on one or more diseases, conditions, or symptoms under treatment. For prophylactic benefit, the compositions may be administered to a subject at risk of developing a particular disease, condition, or symptom, or to a subject reporting one or more of the physiological symptoms of a disease, even though the disease, condition, or symptom may not have yet been manifested.

The phrases "pharmaceutical or pharmacologically acceptable" refers to molecular entities and compositions that do not produce an adverse, allergic, or other untoward reaction when administered to an animal, such as a human, as appropriate. The preparation of a pharmaceutical composition comprising a therapeutic agent, such as a cell, or additional active ingredient will be known to those of skill in the art in light of the present disclosure. Moreover, for animal (*e.g.*, human) administration, it will be understood that preparations should meet sterility, pyrogenicity, general safety, and purity standards as required by FDA Office of Biological Standards. As used herein, "pharmaceutically acceptable carrier" includes any and all aqueous solvents (*e.g.*, water, alcoholic/aqueous solutions, saline solutions, parenteral vehicles, such as sodium chloride, Ringer's dextrose, etc.), non-aqueous solvents (*e.g.*, propylene glycol, polyethylene glycol, vegetable oil, and injectable organic esters, such as ethyloleate), dispersion media, coatings, surfactants, antioxidants, preservatives (*e.g.*, antibacterial or antifungal agents, anti-oxidants, chelating agents, and inert gases), isotonic agents, absorption delaying agents, salts, drugs, drug stabilizers, gels, binders, excipients,

disintegration agents, lubricants, sweetening agents, flavoring agents, dyes, fluid and nutrient replenishers, such like materials and combinations thereof, as would be known to one of ordinary skill in the art. The pH and exact concentration of the various components in a pharmaceutical composition are adjusted according to well-known parameters.

5 As used herein, the term “contacting,” when used in reference to any set of components, includes any process whereby the components to be contacted are mixed into same mixture (for example, are added into the same compartment or solution), and does not necessarily require actual physical contact between the recited components. The recited components can be contacted in any order or any combination (or sub-combination) and can
10 include situations where one or some of the recited components are subsequently removed from the mixture, optionally prior to addition of other recited components. For example, “contacting A with B and C” includes any and all of the following situations: (i) A is mixed with C, then B is added to the mixture; (ii) A and B are mixed into a mixture; B is removed from the mixture, and then C is added to the mixture; and (iii) A is added to a mixture of B and C. “Contacting”
15 a target nucleic acid or a cell with one or more reaction components, such as an Cas protein or guide RNA, includes any or all of the following situations: (i) the target or cell is contacted with a first component of a reaction mixture to create a mixture; then other components of the reaction mixture are added in any order or combination to the mixture; and (ii) the reaction mixture is fully formed prior to mixture with the target or cell.

20 The term “split” or “in split” when referring to the configuration of CRISPR-Effector protein(s) means that the CRISPR and effector polypeptides are separated into two individual polypeptides. In other words, the CRISPR and effector proteins are not fused covalently to form one single polypeptide.

The term “mixture” as used herein, refers to a combination of elements, that are
25 interspersed and not in any particular order. A mixture is heterogeneous and not spatially separable into its different constituents. Examples of mixtures of elements include a number of different elements that are dissolved in the same aqueous solution, or a number of different elements attached to a solid support at random or in no particular order in which the different elements are not spatially distinct. In other words, a mixture is not addressable.

30 As disclosed herein, a number of ranges of values are provided. It is understood that each intervening value, to the tenth of the unit of the lower limit, unless the context clearly dictates otherwise, between the upper and lower limits of that range is also specifically disclosed. Each smaller range between any stated value or intervening value in a stated range

and any other stated or intervening value in that stated range is encompassed within the disclosure. The upper and lower limits of these smaller ranges may independently be included or excluded in the range, and each range where either, neither, or both limits are included in the smaller ranges is also encompassed within the disclosure, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the disclosure. The term “about” generally refers to plus or minus 10% of the indicated number. For example, “about 10%” may indicate a range of 9% to 11%, and “about 20” may mean from 18-22. Other meanings of “about” may be apparent from the context, such as rounding off, so, for example “about 1” may also mean from 0.5 to 1.4.

EXAMPLES

Material and Methods

Cell and cell culture conditions

HEK293T and K562 cells were purchased from ATCC (HEK293T, CRL-3216; K562, CCL-243). HEK293T Cells were grown and maintained at 37°C and 5%CO₂ in Dulbecco’s modified Eagle’s medium (Thermo Fisher Scientific), supplemented with 10% fetal bovine serum, 1x glutamine (Thermo Fisher Scientific), and 1x antibioticantimycotic solution (Thermo Fisher Scientific). K562 cells were grown in RPMI 1640 medium with 10% fetal bovine serum and 1x glutamine.

Transfections

Transfections by electroporation were performed on a Neon Transfection System (Thermo Fisher Scientific), according to the manufacturer’s instructions. Briefly, 2 x10⁵ cells were transfected with around 2µg of a combination of DNA from Cas9/RT, gRNA/magRNA, template RNA, and episomal target plasmid constructs. For experiments from FIG. 17 to FIG. 29, the amount of each expression vector DNA when used in each electroporation was: 250ng of gRNA or magRNA or molar equivalent pegRNA expression vector, 250 ng of RNA template expression vector, 500 ng nCas9(H840A)-RT fusion protein expression vector, 250 ng of MCP-RT fusion expression vector, and 100 ng target episomal target plasmid DNA (or sometime 500ng if targeting an episomal EGFP sequence, as indicated). For experiments in FIG. 30 and after, the amount of DNA used for nCas9(H840A)-RT, MCP-RT fusion expression

vector, and episomal target plasmid DNA remained the same as described above. The expression vectors for the gRNA, pegRNA, magRNA, and RNA template varied and were described in each experiment. If not specified, the amount of expression vectors for magRNA and second nicking gRNA was 250 ng each. The RNA template expression vector is 1000ng, and the pegRNA expression vector was molar equivalent to 1000ng of the RNA template expression vector. Under this most commonly used conditions in FIG. 30-43, the molar ratio of the RNA template in PE system: pegRNA scaffold in PE: RNA template in ME: magRNA scaffold in ME was roughly equal to 1:1:1:0.25.

The U6 expression cassettes of pegRNA, RNA template, magRNA, and gRNA were cloned into the pBS KS(+) vector (2964 bp), resulting in plasmid sizes around 3500 bp in length. As these plasmids also share U6 promoter and transcription terminator, the size differences between pegRNA, gRNA, magRNA, and RNA template expression vectors are usually within 100bp (out of total length of about 3500 bp). The CMV expression cassettes of proteins such as nCas9, nCas9-RT fusion, MCP-RT, MCP-p65 are cloned into pcDNA3.1(+) (5.4kb).

When comparing the activity between ME and PE systems, the molar ratio of the RNA template in PE system (in pegRNA): pegRNA scaffold in PE: RNA template in ME: magRNA scaffold in ME was roughly equal to 1:1:1:0.25 as described above. Student t-test was performed in comparisons, with * indicating $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$.

After electroporation, cells were seeded in 12-well plates. Three days after electroporation, EGFP-containing cells were observed under fluorescent microscope and representative pictures were taken. To quantify percentage of fluorescent EGFP cells, cells were subject to flow cytometry analyses. All experiments were performed with two or three biologically independent replicates. To directly determine base changes at the target sequences, DNA sequences containing the target sites (either EGFP or endogenous genes) were PCR amplified, analyzed by Sanger sequencing, and quantified by the EditR analytical tool. Some samples were further analyzed by next-generation sequencing (NGS). Specific conditions for each experiment are further described in the figure legends.

Designing magRNA and RNA template design for Match Editing

For designing a magRNA and an RNA template for a protein encoding or RNA encoding target, the RNA template is usually the sense sequence. This minimizes the potential formation of double strand RNA in cells. In this scenario, the non-coding strand (or lower

strand) is nicked and extended using the sense RNA template sequence. This is named as Lower Strand Nicking and Extension Mode.

For Lower Strand Nicking and Extension Mode, the following were example steps used to design RNA template and magRNA:

- 5 a. Align the two complementary target DNA strands containing the Sense upper strand DNA (5'-3' orientation, right to left) and its complementary anti-sense lower strand DNA (3'-5' orientation, right to left)
- 10 b. Identify the PAM motif at the lower strand, and identify the nicking site at the lower strand (lower strand is the non-target strand, upper strand is the target strand complementary to the guide or spacer sequence, guide is immediately 5' to PAM)
- 15 c. To design the RNA template, with a length of 13-nucleotide prime binding sequence (PBS) and a 35-nucleotide RT extension length as an example:
 - i. Identify lower strand cutting site (3 nt 5' to PAM);
 - 20 ii. identify UPPER STRAND PBS: starting from THE NUCLEOTIDE (UPPER Strand) COMPLEMENTARY to the cutting nucleotide (LOWER), count around 13 nt towards 3' end the upper strand (ideally ending with a C or G);
 - 25 iii. Upper strand RT Extension: counting 35 nt (ideally ending with a G) towards 5' end of the upper strand, starting from the complementary nucleotide of cutting nucleotide; or depending on the nature of deletion/insertion, choose an appropriate region of homologous sequence for RT extension.
 - 30 iv. Copy the full 5'-RT Extension (35 nt) +PBS (13 nt)-3' sequence from the upper strand to a new file
 - 35 v. Modify the template by adding THE DESIRED MUTATIONS to the RT extension sequence in the above file.
 - 40 vi. The resulting sequence is the Match Editing RNA template
- 45 d. To design the magRNA sequence, using a 12-nt anchoring tag as an example:
 - i. Locate PAM and Guide from lower strand;
 - ii. Copy guide (lower strand, ~20 nt, 5' of PAM, COPY in 5'-3' orientation)
 - iii. Paste guide/spacer sequence to the 5' end of gRNA scaffold (using spCas9 gRNA scaffold as an example)
 - iv. From the template, find a 12- nucleotide sequence to be anchored (*e.g.*, a sequence starting at 29-nt 5' of the PBS sequence)
 - v. copy the Bottom strand (5'-3') that is complementary to the 12-nucleotide sequence
 - vi. AND PASTE it into the 3' end of gRNA (use the spCas9 gRNA scaffold as an example, with the anchoring tag added to the 3' end of gRNA without a linker as an example).
 - vii. The magRNA having the 5'-spacer/guide+gRNA scaffold+3' template anchoring tag sequence of 12 nt.

For designing an RNA template and magRNA with upper strand nicking, example Upper Strand Nicking and Extension Mode design steps were described below:

- a. To design the RNA template, with 13-nucleotide prime binding sequence and 35-nucleotide RT extension length as an example:
 - i. Align the two complementary target DNA strands containing the upper strand DNA (5'-3' orientation, right to left) and the complementary lower strand DNA (3'-5' orientation, right to left)
 - ii. Identify the Upper strand PAM motif and nicking site (upper strand is the non-target strand, lower strand is the target strand containing sequence complementary to the guide/spacer sequence in gRNA/magRNA, guide is immediately 5' to PAM, nicking site is 3 nt 5' to PAM);
 - iii. READ LOWER STRAND: for PBS with 13-nucleotide length in design: starting from THE COMPLEMENTARY NUCLEOTIDE (lower) of the cutting site (Upper), count 13 nt towards 3' end of the lower strand;
 - iv. Lower strand RT Extension length of 35 nt: counting 35 nt towards 5' end of the Lower strand, starting from the complementary nucleotide of cutting site; or depending on the nature of deletion/insertion, choose an appropriate region of homologous sequence for RT extension.
 - v. Copy Bottom sequence containing above PBS and RT extension sequence (copy with orientation of 5'-3') and paste to template plasmid
 - vi. Modify the template by adding THE DESIRED MUTATIONS to the RT extension sequence.
 - vii. The resulting sequence is the Match Editing RNA template
- b. To design a magRNA sequence, using a size of 12-nt anchoring tag as an example:
 - i. Identify the Upper strand PAM motif and nicking site (upper strand is the non-target strand, lower strand is the target strand containing sequence complementary to the guide/spacer sequence in gRNA/magRNA, guide is immediately 5' to PAM);
 - ii. Copy guide (Upper strand, 20 nt, 5' of PAM, COPY in 5'-3' orientation)
 - iii. Paste guide sequence to 5' end of gRNA scaffold (using spCas9 gRNA scaffold as an example)
 - iv. From the template, find a 12- nucleotide sequence to be anchored (*e.g.*, a sequence starting at 29-nt 5' of the PBS sequence)
 - v. copy the Bottom strand sequence complementary to the 12-nucleotide sequence (copy with 5'-3' orientation) to be anchored
 - vi. PASTE it into the 3' end of gRNA (use spCas9 gRNA scaffold as an example, with the anchoring tag added to the 3' end of gRNA without a linker as an example).
 - vii. The magRNA having the 5'-spacer/guide+gRNA scaffold+3' template anchoring tag sequence of 12 nt.

The RNA template or magRNA sequence was inserted to an expression plasmid between the U6 promoter sequence and U6 transcription terminator for expression in cells. It

is important to note that the description is to provide some examples but not to provide limits on how the system are designed.

Sequences

Listed below are exemplary sequences used in the examples.

5 *Nickase spCas9-RT sequence (NLS-spCas9 (H840A)-MMTV-RT(mut)-NLS):*
 MPKKKRKVDKKYSIGLDIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDS
 GETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHP
 IFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHAMIKFRGHFLIEGDLNPDNS
 DVDKLFIQLVQTYNQLFEEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNL
 10 IALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLS
 DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG
 GASQEEFYKFIKPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIPHQIHLGELHAILRRQE
 DFYPFLKDNREKIEKILTFRIPIYVVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQ
 SFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL
 15 LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIKDKDFLDNEENE
 DILEDIVLTTLTFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
 GKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQDSLHEHIANLAGSPAIIKKGI
 LQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEH
 PVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVEDAIVPQSFLKDDSIDNKVLTRSD
 20 KNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTAKERGGLSELKAGFIKRQLVE
 TRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYYHHAH
 DAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVRKMIKSEQIEGKATAKYFFYSNIMNFFK
 TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIIVKKTEVQTGGFSKES
 ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMER
 25 SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYV
 NFLYLASHYEKLKGSPEQKQLFVEQHKHYLDEIIIEQISEFSKRVIILADANLDKVL SAYN
 KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSIITGLYE
 TRIDLSQLGGDSGGSSGGSSGSETPGTSESATPSSGGSSGGSSSTLNIEDEYRLHETSKEPD
 VSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIIKQYPMSQEARLGKPHIQR
 30 LLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNRVEDIHPTVPNPYNLLSGLPPS
 HQWYTVLDLKDFAFFCLRLHPTSQPLFAFEWRDPENGISGQLTWTRLPQGFKNSPTLFNEALH
 RDLADFRIQHPDLIILLQYVDDLLLAATSELDCCQOQTRALLQTLGNLGYRASAKKAQICQKQV
 KYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFI PGFAEMAAPLYPLT
 KPGTLFNWGPDQQKAYQEI KQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGWRR
 35 PVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWL
 SNARMTHYQALLDTRVQFGPVVALNPATLLPLPEEGLQHNCILDILAEAHGTRPDLTDQPL
 PDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGK
 KLVNYTDSRYAFATAHIGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGH
 QKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSPSGGSKRTADGSEFEPKKRKV
 40 (SEQ ID NO: 67)

Nickase saCas9-RT sequence (NLS-saCas9 nickase (N580A)-MMTV-RT(mut)-NLS):
 MPKKKRKVGKRNIIIGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGAR
 RLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRR
 45 GVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSYVK
 EAKQLLKVKQAYHQLDQSFIDTYIDLLETRRTYYEGPGEKSPFGWKDIKEWYEMLMGHCTYF
 PEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYEKFQIIENVFKQKKKPTLKQIAKE

ILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDI
 QEELTNLNSLTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPK
 KVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIVKINAIKKYGLPNDIIIELAREKNSKDAQ
 KMINEMQKRNRQTNERIEEII RTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNP
 5 FNYEVDHIIIPRSVSFDNSFNKVLVKQEEASKKGNRTPFQYLSSSDSKISYETFKKHILNLA
 KGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLRLSYFRVNNLDVKVK
 SINGGFTSFLRRKWKFKKERNKGYKHHADALI IANADFI FKEWKKLDKAKVMENQMFEED
 QAESMPEIETE QEYKEIFITPHQIKHIKDFKDYKYSHRVDKKNRELINDTLYSTRKDDKGN
 TLIVNNLNLGLYDKDNDKLLKLINKSPEKLLMYHHDPTQYQKLKLIMEQYGDENPLYKYEE
 10 TGNYLTKYSKKDNGPVIKKIKYYGNKLNALHDITDDYPNSRNKVVKLSLKPYRFDVYLDNGV
 YKFVTVKNLVDVIKENYEEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGV
 NNDLLNRIEVMIDITYREYLENMNDKRPPRI IKTASKTQSIKKYSTDILGNLYEVKSKKH
 PQIIKKGSGSGSSGSSGSETPGTSESATPESSGSSGSSSTLNI EDEYRLHETSKEPDVSLG
 STWLSDFPQAWAETGGMGLAVRQAPLI IPLKATSTPVS IKQYPMSQEARLG I KPHIQRLLDQ
 15 GILVPCQSPWNTPLLPVKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY
 TVLDLKD AFFCLRLHPTSQPLFAFEWRDP EMGISGQLTWTRLPQGFKNSPTLFNEALHRDLA
 DFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLG
 YLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGT
 LFNWGPDQQKAYQEI KQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLG PWRRPVAY
 20 LSKKLD PVAAGWP PCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNAR
 MTHYQALLLDTRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDAD
 HTWYTDGSSLLQEGQRKAGAAVTTE TEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN
 YTDSRYAFATAHIGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGH
 SAEARGNRMADQAARKAAITETPDTSTLLIENSSPSGGSKRTADGSEFEPKKKRKV (SEQ
 25 ID NO: 68)

Nickase spCas9 in split ME (NLS-spCas9 (H840A)-NLS) sequence

MPKKKRKVDKKYSIGLDIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHS IKKNLIGALLFDS
 30 GETAEATRLKRTARRRYTRKRNRI CYLQEIFSNEMAKVDDSFHRL EESFLVEEDKKHERHP
 IFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNS
 DVDKLFIQLVQTYNQLF EENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNL
 IALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLF LAAKNLSDAILLS
 DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG
 35 GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQE
 DFYPFLKDNREKIEKILTFRIPIYVGPLARGNSRFAMWTRKSEETITPWNFE EVVDKGASAQ
 SFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL
 LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI IKDKDFLDNEENE
 DILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
 40 GKTILDFLKSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA IKKGI
 LQTVKVVDLVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEH
 PVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSD
 KNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVE
 TRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYK VREINNYHHAH
 45 DAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMI AKSEQEIGKATAKYFFYSNIMNFFK
 TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNI VVKTEVQTGGFSKES
 ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELLGITIMER
 SSFEKNPIDFLEAKGYKEVKKDLI IKLPKYSLFEL ENGRKRMLASAGELQKGNELALPSKYV
 NFLYLASHYEKLKGS PEDNEQKQLFVEQHKHYLDEIIEQISEF SKRVILADANLDKVL SAYN

KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYE
TRIDLSQLGGDSGGSGGSGGSPKKKRKV (SEQ ID NO: 69)

5 Nickase SaCas9 in split ME (NLS-saCas9 (N580A)-NLS) sequence

MPKKKRKVGKRNII LGLDIGITSVGYGIIDYETRDVIDAGVRLFK EANVENNEGRRSKRGAR
RLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRR
GVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTS DYVK
EAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYF
10 PEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYEKFQIIENVFKQKKKPTLKQIAKE
ILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEI IENAELLDQIAKILTIYQSSEDI
QEELTNLNSLTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAI FNRLKLVPK
KVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAI IKKYGLPNDII IELAREKNSKDAQ
KMINEMQKRNRQTNERIEEII RTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNP
15 FNYEVDHII PRSVSFDNSFNNKVLVKQEEASKKGNRTPFQYLSSSDSKI SYETFKKHILNLA
KGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNNLLRSYFRVNNLDVKVK
SINGGFTSFLRRKWKFKKERNKGYKHH AEDALI IANADFI FKEWKKLDKAKVMENQMFE EK
QAESMPEIETE QEYKEI FITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGN
TLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHPQTYQKLKLIMEQY GDEKNPLYKY YEE
20 TGNLYTKYSKKDNGPVIKKIKY YGNKLNALHDITDDYPNSRNKVVKLSLKP YRFDVYLDNGV
YKFVTVKNLDVIKENY YEVNSKCYEEAKKLLKISNQA EFIASFYNNDLIKINGELYRVIGV
NNDLLNRIEVNMIDITYREYLENMNDKRPPRI IKTIA SKTQSIKKYSTDILGNLYEVKSKKH
PQIIKKSGGSGGSGGSPKKKRKV (SEQ ID NO: 70)

25

The nickase saCas9 (N580A)-KKH variants sequences are the same as the saCas9 (N580A) sequences, except for three mutations (E782K/N968K/R1015H)

30 NLS-MCP-L25-MMLV(mut)-RT-NLS sequence:

PKKKRKVTLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLI IPLKAT
STPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLRE
VNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKD AFFCLR LHPTSQPLFAFEWRDPEMGI
SGQLTWTRL PQGFKNSPTLFNEALHRDLADFR IQHPDLILLQYVDDLLLAATSELDCQQGTR
35 ALLQTLGNLGYRASAKKAQICQKQVKYLG YLLKEGQRWLTEARKETVMGQPTPKTPRQLREF
LGKAGFCRLFI PGFAEMAAPLYPLTKPGTLFNWGPDQOKAYQEIKQALLTAPALGLPDLTKP
FELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAI AVLTKDAGKLT
MGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLD TDRVQFGPVVALNPATLLPLPEE
GLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKA
40 LPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKN
KDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPD TSTLLIENS
SPSGGSKRTADGSEFEELKTPLGDTTHTSPPCPAPELLGGPMASNFTQFVLVDNGGTGDVTV
APSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLN MELTIPIF
ATNSDCELI VKAMQGLLDGNNPIPSAIAANS GIYEFEPKKKRKV (SEQ ID NO: 71)

45

Note:

In the examples above, Nuclear localization signals (NLS) are added at the N-terminus or C-terminus of the fusion proteins, and linker peptides are added between the fusion moieties.

5 NLS-MCP-L25-p65-NLS sequence

PKKKRKVSNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNR
KYTIKVEVPKGAWRSYLNMEITPIPIFATNSDCELVKAMQGLLKDGNIPIPSAIAANSIYEL
KTPLGDTTHTSPPCPAPELLGGPYLPDTPDRHRIEEKRKRTYETFKSIMKKSPPFSGPTDPRP
PPRRIAVPSRSSASVPKPAPQYPFTSSLSTINYDEFPTMVFPSPGQISQASALAPAPPQVLP
10 QAPAPAPAPAMVSALAQAPAPVPVLAPGPPQAVAPPAPKPTQAGEGTLSEALLQLQFDDDL
GALLGNSTDPVFTDLASVDNSEFQQLLNQGI PVAPHTTEPMLMEYPEAITRLVTGAQRPPD
PAPAPLGAPGLPNGLLSGDEDFSSIADMDFSALLSQTRPKKKRKVEQKLISEEDL (SEQ ID
NO: 72)

15

wtEGFP and mutant EGFP genes:

ATGgtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcg
Δ100G
acgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagct
20 Δ50C Δ35C Δ20C
gaccctgaagttcatctgcacacccggcaagctgcccggtgccctggcccaccctcgtagaccacc
Δ4A A200
ctgacctaaccggcgtgcagtgccttcagccgctaccccgaccacatgaagcagcagcacttcttcaagtc
cgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagaccc
25 ggcggaggtgaagttcgagggcgacaccctggtgaaccgcacatcgagctgaagggcatcgacttcaag
gaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggc
cgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgc
agctcgccgaccactaccagcagaacacccccatcgggcgacggccccgtgctgctgcccgacaaccac
tacctgagcaccacagtcggccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctgga
30 gttcgtgaccgcccgggatcactctcggcacatggacgagctgtacaag (SEQ ID NO: 73)

Note:

- a. The sequence encodes a wild type EGFP.
- 35 b. nfEGFP contains a A 200 G point mutation (A200G, Tyr66Cys). A200 position in the wild type EGFP gene is marked and underlined.
- c. The nucleotides deleted in the deletion constructs, EGFP Δ4A, Δ 20C, Δ 35C, Δ 50C, and Δ1 00G, are labeled and marked (underlining). The relative numbers 4, 20, 35, 50, and 100, are relative distance towards the lower strand NGG PAM (where N is
- 40 labeled as +1, corresponding to T199 position).
- d. The EGFP insertion construct, EGFP-4nt-insertion, contains a four-nucleotide (ATAG) between nucleotides 179 and 180.

- e. EGFP123Δ52nt contains a deletion from nucleotide 124- to 175 (a 52-nucleotide deletion).

5 *U6 promoter and terminator sequence:*

GAGGgcctattttcccatgattccttcatatttgcatatacgatacaaggctgttagagagat
aattggaattaaatttgactgtaaacacaaagatattagtacaaaaatacgtgacgtagaaagt
aataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcatatgctt
accgtaacttgaaagtattttcgatttcttggctttatatcttgTGGAAAGGACGAAACAC
 10 C TTTTCTCCGCTGAGCGTACTGAGACGCC (SEQ ID NO: 74)

Note:

The single underlined sequence is the U6 promoter sequence, the double underlined sequence
 15 is used as U6 terminator. Template RNA, gRNA, or magRNA sequence is inserted between
 the promoter and terminator sequences for expression.

U6, ribozyme HDV, U6 Terminator:

GAGGgcctat

20 aattggaatta

aatttgactg

taaacacaaagata

ttagta

caaaaatacgtgacg

tagaaagt

aataatttcttgggtag

tttgcag

ttttaaaattatg

ttttaaaatggactat

catatgctt

accgtaacttgaaagt

at

tttcgatttcttggctt

tatatctt

gTGGAAAGGACGAAACAC

CGGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTTCGGCATGGCGA

25 ATGGGACTTTTCTCCGCTGAGCGTACTGAGACGCC (SEQ ID NO: 75)

Note:

The underlined sequence is ribozyme HDV sequence. The magRNA is inserted between the
 U6 promoter and the HDV sequence.

HEK4 genomic sequence with Guide Underlined and PAM double underlined:

CGGCTTCTCCCTCAGTCAGTCCATGCCTGCAGGGTCTGGAACCCAGGTAGCCAGAGACCCGC

TGGTCTTCTTTCCCTCCCTGCCCTCCCTCCCTTCAAGATGGCTGACAAAGGCCGGGCTG

GGTGAAGGAAGGGAGGAAGGGCGAGGCAGAGGGTCCAAAGCAGGATGACAGGCAGGGGCAC

35 CGCGGCGCCCCGGTGGCACTGCGGCTGGAGGTGGGGGTAAAGCGGAGACTCTGGTGCTGTG

TGACTACAGTGGGGGCCCTGCCCTCTCTGAGCCCCCGCCTCCAGGCCTGTGTGTGTGTCTCC

GTTCCGGTTGAAAGGAGCCCGGGAAGGCCCCAGAGGAGTCTGGTTTTGGACGTCTGAC

CCCACCCCTCCCGCTTAGGGCTTCTGATCCCCAGGGTGATTTCACTGGCC (SEQ ID NO:

76)

HEK3 genomic sequence with Guide Underlined and PAM double underlined:

CTTGGCATGAGAAACCTTGAGAGTTTTAAAGCAAGGGCTGATGTGGGCTGCCTAGAAAGGCA

TGGATGAGAGAAGCCTGGAGACAGGGATCCCAGGGAAACGCCCATGCAATTAGTCTATTTCT

45 GCTGCAAGTAAGCATGCATTTGTAGGCTTGATGCTTTTTTCTGCTTCTCCAGCCCTGGCCT

GGGTCAATCCTTGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAGCCCTGCTTCCT
CCAGAGGGCGTCGCAGGACAGCTTTTCCCTAGACAGGGGCTAGTATGTGCAGCTCCTGCACCG
GGATACTGGTTGACAAGTTTGGCTGGGCTGGAAGCCAGCACCTAGGGAGGTCCCTGGAAGGG
GCCAGCCTCACCAGGAGAGGAGGGACCTGGCCCTTCAGGGTCGAGCTCAAC (SEQ ID NO:

5 77)

Splicing reporter containing 5' half EGFP-artificial intron-3' half EGFP:

ATGGTGAGCAAGGGCGAGGAGCTGTTTACCGGAGTTGTCCCCATTCTGGTCGAGCTTGATGG
CGACGTGAATGGACACAAGTTTAGCGTaaagtGGAGAAGGAGAAGGAGATGCCACgTATGGGA
10 AACTGACCCTGAAGTTCATCTGCACAACCTGGGAAATTGCCTGTGCCTTGCCCGACTtTgGTG
ACAACCCCTCACTTATGGCGTGCAGTGTTCCTCCaGATACCCCGATCACATGAAACAGCATGA
CTTCTTTAAGTCTGCCATGCCAGAAGGgTATGTACAGGAACGTACCATCTTCTTCAAAGATG
ACGGCAACTACAAGACGAGGGCTGAgGTGAAGTTTGAGGGtGACACACTCGTCAATCGGATT
GAGCTGAAAGGCATCGACTTTAAGGAGGATGGGAACATACTTGGTCACAAGCTGGAgTACGT
15 AGGTTATGCATTAATTTTTATATCTGTACTCATTTTGTGCTGCTTGTAAGTCCGTGCTTTG
TTAAAGCTTGGTACCGAGCTCGGATCCACTAGTCCAGTGTGGTGAATTTCGCCCTTAAGGGC
GAATTCTGCAGATATCTGGTTTGAAATCATTCATGTGGTGAATACTTTTATAATTAGGATGT
GTTGGCTTTCAGATCATTTCTTTCAGTCTGTGGGTTTCAGGGGATATATTTAATTATTTTTTT
CTTTCTAGAACTACAACCTCCATAACGTGTATATCATGGCTGACAAGCAGAAGAATGGCATA
20 AAgGTGAACCTCAAAATCCGCCATAATATCGAGGATGGTTCAGTCCAACCTCGCAGACCATTA
CCAGCAGAACACTCCAATTGGCGATGGACCTGTGTTGCTTCCaGACAATCACTACCTGTCAA
CCCAATCTGCGCTGAGCAAAGACCCCAATGAGAAACGAGATCACATGGTCTGCTGGAGTTT
GTGACAGCAGCCGGGATTACCTCGGCATGGACGAACCTGTATAA (SEQ ID NO: 78)

25 Note:

The beginning first single underlined sequence is 5' half GFP sequence, the ending single underlined sequence is the 3' half GFP sequence. The sequence in between is an artificial intron with SAS and SDS sites at the intron-exon junctures.

30 *Mouse Dmd exon 23 skipping reporter containing 5' half EGFP-Dmd Intron 22-Dmd Exon 23-Dmd Intron 23-3' half EGFP reporter sequence:*

ATGGTGAGCAAGGGCGAGGAGCTGTTTACCGGAGTTGTCCCCATTCTGGTCGAGCTTGATGG
CGACGTGAATGGACACAAGTTTAGCGTaaagtGGAGAAGGAGAAGGAGATGCCACgTATGGGA
AACTGACCCTGAAGTTCATCTGCACAACCTGGGAAATTGCCTGTGCCTTGCCCGACTtTgGTG
35 ACAACCCCTCACTTATGGCGTGCAGTGTTCCTCCaGATACCCCGATCACATGAAACAGCATGA
CTTCTTTAAGTCTGCCATGCCAGAAGGgTATGTACAGGAACGTACCATCTTCTTCAAAGATG
ACGGCAACTACAAGACGAGGGCTGAgGTGAAGTTTGAGGGtGACACACTCGTCAATCGGATT
GAGCTGAAAGGCATCGACTTTAAGGAGGATGGGAACATACTTGGTCACAAGCTGGAgTACGT
AGGTTATGCATTAATTTTTATATCTGTACTCATTTTGTGCTGCTTGTAAGTCCGTGCTTTG
40 TTAAAGCTTGGTACCGAGCTCGGATCCACTAGTCCAGTGTGGTGAATTTCGCCCTTttctaa
atthctgtctaaatataatagccctgtaataatagaaattattcataagaatacatata
tattgctttatcagatatthctactttgtttagatctctaaattacataaactttttatttacc
ttctttcttgatatgaatgaaactcatcaaatatgctgttagtgtaaataaacttctattta
atthttgagGCTCTGCAAAGTTCTTTGAAAGAGCAACAAAATGGCTTCAACTATCTGAGTGAC
45 ACTGTGAAGGAGATGGCCAAGAAAGCACCTTCAGAAATATGCCAGAAATATCTGTGAGAATT
TGAAGAGATTGAGGGGCACTGGAAGAACTTTCTCCAGTTGGTGGAAAGCTGCCAAAAGC
TAGAAGAACATATGAATAAACTTCGAAAATTTCAAGgtaagccgaggtttggccttttaaacta
tattttttcacatagcaattaattggaaaatgtgatgggaaacagatatthttaccagagtc

cttcaaagatatattgatgatataaaaagccaaatctatcttcaaaggattgcaacttgccatt
 tttcctatgaaaacagtaatgtgtcataccttcttggaattgtctgtataAAGGGCGAATTCT
 GCAGATATCTGGTTTGAAATCATTCATGTGGTGAATACTTTTATAATTAGGATGTGTTGGCT
 TTCAGATCATTTCTTTTCTAGTCTGTGGGTTTCAAGGGATATATTTAATTATTTTTTCTTTCTA
 5 GAACTACAACCTCCCATAACGTGTATATCATGGCTGACAAGCAGAAGAATGGCATAAAgGTGA
ACTTCAAAAATCCGCCATAATATCGAGGATGGTTCAGTCCAACCTCGCAGACCATTACCAGCAG
AACACTCCAATTGGCGATGGACCTGTGTTGCTTCCaGACAATCACTACCTGTCAACCCAATC
TGCGCTGAGCAAAGACCCCAATGAGAAACGAGATCACATGGTTCTGCTGGAGTTTGTGACAG
CAGCCGGGATTACCCTCGGCATGGACGAACGTGTATAA (SEQ ID NO: 79)

Note:

The beginning first single underlined sequence is 5' half GFP sequence, the ending single underlined sequence is the 3' half GFP sequence. The double underlined sequence is the mouse Dmd exon 23 sequence. The sequence between 5' half GFP and exon 23 sequences contains mouse Dmd intron 22, while the sequence between exon 23 and 3' half GFP sequences contains Dmd intron 23.

EGFP ME template	Template sequence	SEQ ID NO:
Ex29-P14	gcccaccctcgtgaccaccctgacc tAcggcgtgcagtgtcttc	80
Ex57-P14	gcaccaccggcaagctgcccgtgcc ctggcccaccctcgtgaccaccctg acctAcggcgtgcagtgtcttc	81
Ex111-P14	ccggcgagggcgagggcgatgccac ctacggcaagctgaccctgaagtgc atctgcaccaccggcaagctgcccg tgccctggcccaccctcgtgaccac cctgacctAcggcgtgcagtgtcttc	82
Ex122-P14	gttcagcgtgtccggcgagggcgag ggcgatgccacctacggcaagctga ccctgaagttcatctgcaccaccgg caagctgcccgtgcccctggcccacc ctcgtgaccaccctgacctAcggcg tgagtgtcttc	83
EGFP gRNA or magRNA Guide	Sequence	SEQ ID NO
nfEGFP-200L	gctgaagcactgcacgccgC	84
wtEGFP-200L	gctgaagcactgcacgccgT	85
EGFP-92U	Ggccacaagttcagcgtgtc	86

EGFP-119U	ggcgagggcgatgccaccta	87
EGFP-151U	ctgaagttcatctgcacCac	88
EGFPΔ50C-151-up	Gcctgaagttcatctgcacac	89

pegRNA RT extension and PBS sequences for EGFP Guide (gctgaagcactgcacgccgC/T)		SEQ ID NO
name	3' RT extension and PBS sequences	
pegEx13-P14	5' accctgacctAcggcgtgcagtgccttc-3'	90
pegEx20-P14	5' cgtgaccaccctgacctAcggcgtgcagtgccttc-3'	91
pegEx35-P14	5' gccctggcccaccctcgtgaccaccctgacctAcggcgtgcagtgccttc-3'	92
	5' cggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctAcggcGtgcagtgccttc-3'	93
pegEx109-P14	5' ccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatCtgcaccaccggcaagctgcccgtgccctggcccaacctcgtgaccaccctgacctAcggcgtgcagtgccttc-3'	94

magRNA match sequences for EGFP-200L Guide are previously listed in the Matching gRNA examples section.

5

Some additional magRNA information is listed below:

magRNA for Site HEK3 (Site 3)		
Name	5' Guide	3' match sequence
13M-Ex25	5' GGCCCAGACTGAGCACGTGA-3' (SEQ ID NO: 95)	5' - AAGGAAGCCCTG -3' (SEQ ID NO: 96)
magRNA for Site HEK4 (Site 4)		
Name	5' Guide	3' match sequence
12M-Ex24	5' GGCAGTGCAGGCTGGAGGTGG-3' (SEQ ID NO: 97)	5' - AGACTCTAGTGC -3' (SEQ ID NO: 98)
RNA Template for HEK3 and HEK 4		
HEK3 template-Ex25-2GT-9GC-P13	5' -CAGGGCTTCCTTTgCTCTGCAATCACGTGCTCAGTCTG-3' (SEQ ID NO: 99)	
HEK4 template-Ex24-2GA 20GA	5' - GCACTAGAGTCTCCGCTTTAACTCCACCTCCAGCCGC-3' (SEQ ID NO: 100)	

mDmd exon 23/intron 23 SDS deletion template (DNA encoding the RNA sequence)

Low_mEx23-RT58del29-P14	GCTGCCAAAAGCTAGAGAACATATgagggttggcctttaaac (SEQ ID NO: 101)
mDmd exon 23/intron 23 SDS deletion pegRNA 3'-sequence:	5' - gctatgtgaaaaaatatagtttaaaggccaaaaattaaatagaag ttc-3' (SEQ ID NO: 102)

Guide and Mag tag sequences for mDmd Exon 23/intron 23 SDS/gRNA:		
Name	5' Guide sequence	3' Mag tag sequence
magLow	Gatagtttaaaggccaaacct (SEQ ID NO: 103)	GCTTTTGGCAGC (SEQ ID NO: 104)
Sg RNA nicking	GAAACTTTCCTCCCAGTTGG (SEQ ID NO: 105)	

HEK3 (Site3)

Template Name	Sequence
Site3_T25-P13-5GT-12GC	CAGGGCTTCCTTTgCTCTGCaATCACGTGCTCAGTCTG (SEQ ID NO: 106)
Site3_T43-P13-5GT-12GC	GACGCCCTCTGGAGGAAGCAGGGCTTCCTTTgCTCTGCaATCACGTGCTCAGTCTG (SEQ ID NO: 107)
Site3_T43-P13-5GT-12GC-18AC-24TC-30CT	GACGCCCTCTGGAaGAAGCgGGGCTgCCTTTgCTCTGCaATCACGTGCTCAGTCTG (SEQ ID NO: 108)
Site3_T46-P13-5CTTins	GACGCCCTCTGGAGGAAGCAGGGCTTCCTTTCCTCTGCaagCATCACGTGCTCAGTCTG (SEQ ID NO: 109)
Site3_T50-P13-5AP1ins	GACGCCCTCTGGAGGAAGCAGGGCTTCCTTTCCTCTGctgactcaCATCACTGCTCAGTCTG (SEQ ID NO: 110)
Site3_T79-P13-5GILins	GACGCCCTCTGGAGGAAGCAGGGCTTCCTTTCCTCTGCgATcAGgTTgCCcTGgCTgCCgCCgGTcACgGCCATCACGTGCTCAGTCTG (SEQ ID NO: 111)
Site3_T40-P13-6GCAdel	GACGCCCTCTGGAGGAAGCAGGGCTTCCTTTCCTCCATCACGTGCTCAGTCTG (SEQ ID NO: 112)

gRNA/magRNA/pegRNA (SpCas9) name	Guide	Matching Tag (magRNA) / Template+PBS (pegRNA)
phU6_Site3_0xMS2	GGCCCAGACTGAGCA CGTGA (SEQ ID NO: 113)	
magSite3_12M25	GGCCCAGACTGAGCA CGTGA (SEQ ID NO: 114)	AAGGAAGCCCTG (SEQ ID NO: 115)

magSite3_12M42	GGCCCAGACTGAGCA CGTGA (SEQ ID NO: 116)	TCCAGAGGGCGT (SEQ ID NO: 117)
pegSite3_RT24- P13-5GT-12GC	GGCCCAGACTGAGCA CGTGA (SEQ ID NO: 118)	AGGGCTTCCTTTgCTCTGCaATCACGTGCTC AGTCTG (SEQ ID NO: 119)
pegSite3_RT43- P13-5GT-12GC	GGCCCAGACTGAGCA CGTGA (SEQ ID NO: 120)	GACGCCCTCTGGAGGAAGCAGGGCTTCCTTT gCTCTGCaATCACGTGCTCAGTCTG (SEQ ID NO: 121)
pegSite3_RT43- P13-5GT-12GC- 18AC-24TC-30CT	GGCCCAGACTGAGCA CGTGA (SEQ ID NO: 122)	GACGCCCTCTGGAAaGAAGCgGGGCTgCCTTT gCTCTGCaATCACGTGCTCAGTCTG (SEQ ID NO: 123)
pegSite3_RT46- P13-5CTTins	GGCCCAGACTGAGCA CGTGA (SEQ ID NO: 124)	GACGCCCTCTGGAGGAAGCAGGGCTTCCTTT CCTCTGCaagCATCACGTGCTCAGTCTG (SEQ ID NO: 125)
pegSite3_RT50- P13-5AP1ins	GGCCCAGACTGAGCA CGTGA (SEQ ID NO: 126)	GACGCCCTCTGGAGGAAGCAGGGCTTCCTTT CCTCTGctgactcaCATCACGTGCTCAGTCT G (SEQ ID NO: 127)
pegSite3_RT79- P13-5GILins	GGCCCAGACTGAGCA CGTGA (SEQ ID NO: 128)	GACGCCCTCTGGAGGAAGCAGGGCTTCCTTT CCTCTGCgATcAGgTTgCCcTGgCTgCCgCC gCCgGTcACgGCCATCACGTGCTCAGTCTG (SEQ ID NO: 129)
pegSite3_RT40- P13-6GCAdel	GGCCCAGACTGAGCA CGTGA (SEQ ID NO: 130)	GACGCCCTCTGGAGGAAGCAGGGCTTCCTTT CCTCCATCACGTGCTCAGTCTG (SEQ ID NO: 131)
pLow_Site3+63_ 0xMS2	GCACATACTAGCCCC TGTCT (SEQ ID NO: 132)	
pLow_Site3+63_ SLMS2	GCACATACTAGCCCC TGTCT (SEQ ID NO: 133)	

EGFP

Template name	Sequence	
LowEGFP_T51-P13-6CT- 38Gins	gtgtccggcgagGgcgagggcgatgccacctacggcaagctg accTtgaagttcatctgcacca (SEQ ID NO: 134)	
Name of gRNA/magRNA/ pegRNA (SaCas9) name	Guide sequence	Matching Tag (magRNA) / Template+PBS (pegRNA)
LowEGFP_SagRNA_0xMS2	Gccggtggtgcaga tgaactt (SEQ ID NO: 135)	
magLow_EGFP_11M51	Gccggtggtgcaga tgaactt (SEQ ID NO: 135)	Tcgccggacac (SEQ ID NO: 136)

pegLow_EGFP_T51-P13-6CT-38ins	Gccggtggtgcaga tgaactt (SEQ ID NO: 135)	gtgtccggcgagGgcgagggcga tgccacctacggcaagctgaccT tgaagttcatctgcacca (SEQ ID NO: 137)
EGFP_SagRNA+111_0xMS2	Gcaagggcgaggag ctgttcac (SEQ ID NO: 135)	
EGFP_SagRNA+111_SLMS2	Gcaagggcgaggag ctgttcac (SEQ ID NO: 135)	

HBB

template name	Sequence	
HBB_T44-P13-5GT	5' – gttcaccttgccccacagggcagtaacggcagacttctcAAcagga gtcagatgcac-3' (SEQ ID NO: 138)	
mag/gRNA (spCas9) name	guide	Matching Tag (magRNA)
magHBB_Up-12M44	5' – Gcatggtgcatctgactcctg- 3' (SEQ ID NO: 139)	5' –ggcaagggtgaac-3' (SEQ ID NO: 140)
phU6_HBB_gRNA_0xMS2	5' – Gcatggtgcatctgactcctg- 3' (SEQ ID NO: 141)	
pLow_HBB+72	5' – Gccttgataccaacctgccca- 3' (SEQ ID NO: 142)	

5

Example 1

This example describes a specific Match Editing (ME) system that targets an EGFP with a point mutation at position 200 (A200G). The system was delivered by plasmids containing a U6 promoter for the transcription of a RNA template, a magRNA, and a control gRNA (FIG. 17A), and a CMV promoter for nCas9-RT fusion protein and EGFP expression (FIGs. 17A and 17B). The target site of EGFP 200L is also listed and dissected in the figures, where the LOW strand PAM (3'-gga-5') is underlined followed by the guide sequence. Moreover, the mutant base pair, nicking site, and the primer binding sequence are all indicated in FIG. 17C. In addition, the RNA template design is illustrated in FIG. 17D., which contains a 3'-primer binding site (P14) that is exactly the same as the sequence illustrated in the target site, and a 5' RT extension, that is exactly the same as the sequence 5' to the primer binding site in the target site, except for the nucleotide(s) to be changes. In the experiment for correcting

15

the point mutation of A200G, two templates each with a 5' RT extension of 29 or 57 nucleotides were used.

Example 2

5 This example describes the experimental results using the ME system described in Example 1 to correct a point mutation.

Briefly, cells were electroporated with plasmids expressing indicated ME components or control components. The percentage of cells expressing fluorescent EGFP was determined by flow cytometry (FIG. 18A), and the target EGFP DNA fragments were amplified and the
10 complementary strands were sequenced from control (FIG. 18B) or ME (FIG. 18C) treated cells. The results are shown in FIG. 18A (Lane 1, untreated cells; Lane 2, nfEGFP expressing cells; Lane 3-6, cells expressing various ME components as indicated; Lane 7, Prime Editing control). FIGs. 18B and 18C show sequencing results from Lane 2 and Lane 4, respectively.

As shown in FIG. 18A, ME with a magRNA exhibited high gene edition efficiency, a
15 60% of functional editing efficiency for a mutation in EGFP (Lane 4). Moreover, comparing to the gRNA without the match tag complementary to the template, magRNA drastically increased editing efficacy (over 3-fold increase, Lane 3 and 4). The efficacy was similar to the Prime Editing counterpart (no statistical difference between Lane 4 and Lane 7).

FIGs. 18B and 18C showed the confirmation that the conversion of a non-fluorescent
20 EGFP to fluorescent EGFP was indeed caused by point mutation of the complementary C to T. Importantly, the sequencing results (FIG. 18C) clearly demonstrated absence of bystander editing, *i.e.*, only the target C was converted to T but not the Cs close by. Lack of bystander editing is one unique advantage over base editing. The discrepancy in editing efficiency of 26% by direct Sanger sequencing vs 60% by functional EGFP assay of sorting fluorescent cells was
25 due to the fact that each cell had multiple copies of the EGFP genes. The results clearly demonstrated functionality, utility, and effectiveness of the technology of Match Editing by anchoring the RNA template with a complementary matching tag added to the magRNA.

Example 3

30 This example compared the editing efficacy between a long RNA template (Ex57-P14) and a short RNA template (Ex29-P14) in correcting the A200G mutation (3 nucleotide downstream the cutting site).

Cells were electroporated with plasmids expressing indicated ME components or control components, the cells were examined by fluorescent microscopy. The results are shown in FIG. 19A (showing fluorescent field and bright field). The percentage of cells expressing fluorescent EGFP was determined by flow cytometry. The results are shown in FIG. 19B (Panel 1, untreated cells; Panel 2, nfEGFP expressing cells; Panels 4, 6, and 7, Cells expressing ME components with different templates and magRNA; Lane 3 and Lane 5, cells expressing gRNA instead of magRNA).

As shown in the figures, while both templates were efficacious, the shorter one consistently showed higher efficiency (Panel 4 vs. Panel 6-7 in FIGs. 19A and 19B) in correcting this point mutation. Moreover, the efficiency of two different magRNAs that matched either the middle (12M29) of the template or the 5' - end (12M57) of the long template (Ex57-P14) was compared. The data showed that the magRNA matching an internal sequence (12M29) had a higher efficacy than the magRNA matching the distal 5' end sequence (12M57) (Panel 7 vs. Panel 6). In most of the following experiments, the magRNA 12M29, which matched an internal region of the RNA template, was chosen.

Example 4

This example showed the efficacy of ME in correcting deletion mutations by inserting the correct base pair.

Briefly, cells were electroporated with plasmids expressing indicated ME components (12M29 magRNA was used) or control components and examined by fluorescent microscopy. The percentage of cells expressing fluorescent EGFP was determined by flow cytometry. The results are shown in FIGs. 20A and 20B. The target EGFP DNA fragments were amplified and the complementary strands were sequenced from ME treated cells. DNA sequencing results from Panel 3 cells are shown in FIG. 20C.

As shown in the figures, using various correcting RNA templates and the 12M29 magRNA, ME effectively corrected deletion mutations at 4, 20, 35 nucleotides downstream of the nicking site with a template that covers the distance (Panel 3, 4, and 5 in FIGs. 20A and 20B). The correction of the deletion mutation was confirmed by direct Sanger sequencing (FIG. 20C). This study demonstrated that the ME system can be used for insertion editing.

Example 5

This example compared the efficacy of two magRNAs in correcting a deletion at position 35 or 50 downstream the nicking site with a long RNA template (Ex57-P14).

Cells were electroporated with plasmids expressing ME or control plasmids, the percentage of cells expressing fluorescent EGFP was determined by flow cytometry. The results are shown in FIG. 21. It was found that, in both cases, the magRNA matching a proximal sequence of the RNA template (closer to the 3' priming sequence, 12M29) exhibited higher editing efficacy than the magRNA matching the distal 5'-terminal of the RNA template (12M57). In other experiments, the magRNA 12M29 was used in most of the deletion/insertion editing experiments too.

Example 6

This example showed the efficacy of ME in correcting a deletion mutation that is 100 base pair downstream the nicking site. In this experiment, an RNA template with 111-nt long extension sequence (Ex111-P14) and the 12M29 magRNA were used.

Cells were electroporated with an EGFP Δ 100G plasmid and ME plasmids. The target EGFP DNA fragment was amplified and the complementary strand was sequenced from ME treated cells. The results are shown in FIG. 22. The results clearly showed the effectiveness of ME in editing mutations that is distant to the target site, using an independent long RNA template.

Example 7

This example compared the influence of the matching tag length of magRNA and matching position on the template on ME editing efficiency.

Briefly, cells were electroporated with plasmids expressing ME or control plasmids, the percentage of cells expressing fluorescent EGFP was determined by flow cytometry. The results are shown in FIGs. 23A and 23B.

As indicated in FIG. 23A, cells were electroporated the EGFP deletion plasmid Δ 50C alone (EGFP Δ 50), or with EGFP Δ 50 plus ME system differing in magRNA with a matching length varying from 9-nucleotide (9M) to 24-nucleotide (24M), with a 3-nt increment.

As indicated in FIG. 23B, cells were either untreated (NT), or electroporated the EGFP deletion plasmid Δ 35C alone (EGFP Δ 35), or with EGFP Δ 35 plus a ME system differing in magRNA with a matching position in the template, varying in the first matching nucleotide at

position 23, 29, 35, or 41 from the RT extension site (12M23, 12M29, 12M35, and 12M41, respectively). The matching tag length remained constant at 12 nt. length with varying matching the complementary position. The ME template used in both studies was Ex57-P14.

It was found that, by varying the matching length from 9-nucleotide to 24-nucleotide, the matching tag of 9, 12, and 15 exhibited the highest efficacy, while longer matching tags of 18, 21, and 24 nucleotides in length had drastically reduced efficacy. See FIG. 23A.

It was also found that, by varying the matching position from +23 from the priming initiation site (matching sequence between +12 to +23, 12-nt), +29, +35, to +41), the maximal efficacy was achieved with the matching position in the template that was +29 upstream of extension initiation position (matching the template from +18 to +29 from the first template extension nucleotide). See FIG. 23B.

Example 8

This example showed the components and efficacy of a configuration of ME wherein the reverse transcriptase was not directly fused with the RNA-guided nickase, but provided *in split* through recruiting the MCP-RT fusion protein by the MS2 aptamer at the magRNA Stem Loop. The experimental designs and results are shown in FIGs. 24A-24C.

As shown in FIG. 24A, in the split RT ME system, the reverse transcriptase was not fused with nCas9. The magRNA contained an MS2 aptamer at the Stem Loop position. The RT was fused to the MCP through a linker peptide. As shown in FIG. 24B, the gene to be edited was an EGFP gene containing a C deletion 50 nucleotides upstream of the 200-L nicking site. Cells were electroporated with plasmids expressing split RT ME or Fusion RT ME, or with EGFP Δ50 alone, or non-treated, as indicated. The percentage of cells expressing fluorescent EGFP was determined by flow cytometry. The ME template was Ex57-P14; match tag was 12M29.

As shown in FIG. 24C, the ME exhibited lower but comparable efficacy as the direct fusion ME in correcting the deletion at 50 bp downstream the PAM.

Example 9

This example showed a second ME module which only provided a nicking at trans position downstream the first ME module target site drastically enhanced the editing efficacy of the first ME module. The experimental designs and results are shown in FIGs. 25A-25C.

FIG. 25A shows a specific dual module ME system, which includes one template (Ex57-P14), the first magRNA of the first ME module targeting the first target site, the second gRNA of the second ME module targeting the second target site but without a match tag (hence gRNA not magRNA). Functionally, the second ME module only provides a second nicking.

Two second target sites 119-U and 151-U (FIG. 25B) were tested for the first target site 200-L, which are either around 80-nt or 50-nt downstream the first target site.

Cells were electroporated with plasmids expressing a single-module or dual ME, or with EGFP Δ 50 alone, or non-treated, as indicated. The percentage of cells expressing fluorescent EGFP was determined by flow cytometry. The ME template was Ex57-P14, and the match tag was 12M29. As shown in FIG. 25C, the results clearly demonstrated that the second nicking module drastically increases editing efficacy, with the module 80-nt downstream exhibiting better effectiveness.

Example 10

This example showed that Match Editing is effective in editing endogenous gene locus HEK4 site. The experimental designs and results are shown in FIGs. 26A-C.

Briefly, cells were either untreated (FIG. 26A) or electroporated with an ME system with the target site of HEK4 containing either the gRNA without matching tag (FIG. 26B), or a magRNA with appropriate matching tag to template (FIG. 26C). Three days after electroporation, genomic DNA was extracted. The PCR amplified target fragments were sequenced and mutation quantified by EditR program.

As shown in FIGs. 26A-C, using magRNA targeting HEK4 and a template that intends to alter a G to A at +5 position (upstream nicking site), ME led to 10% G to A base change (FIG. 26C). As a control, when gRNA was used instead of magRNA, no editing was observed (FIG. 26B). FIG. 26A showed the non-treated cell HEK4 site sequence. In the figures, the arrows point to editing rate in percentage. It is important to note that for endogenous gene editing, the editing with a gRNA (lack of a matching tag) exhibited no editing effect, in contrast, ME with magRNA exhibited 10% of editing efficiency, indicating the key functional role of a matching tag for Match Editing.

Example 11

This example showed that ME was effective in deletion editing.

An EGFP reporter gene containing 4 nucleotide frameshift insertion starting at position 180 was generated for this experiment. Cells were either untreated, electroporated with expression vector of EGFP with a 4 nucleotide frame-shifting insertion at position 180 (EGFP 180 Ins 4) only, or with the EGFP 180 Ins 4 plus Match Editor (Ex29M12, with 12M29 magRNA and Ex29-P14 template), or with a Prime editor (peg_Ex35). The percentage of fluorescent cells were determined by flow cytometry. The results are shown in FIG. 27.

As shown in FIG. 27, when cells were electroporated with the EGFP insertion construct, only a minimal fluorescent GFP level was observed due to the frameshift mutation. In contrast, when the construct was co-electroporated into the cells with either a ME system or a prime editing system, removal of the frameshift deletion and detection of high levels of EGFP expressing cells were detected. These results demonstrated that ME was effective in deletion editing.

Example 12

This example showed that a single-ME and a dual-ME system were both effective in deleting a splicing regulatory sequence, the splicing donor site (SDS), leading exon skipping. The dual-ME system with the magRNA-gRNA pairing is more efficacious than single-ME system. The experimental designs and results are shown in FIGs. 28A and B.

A splicing reporter construct was generated by splitting the EGFP coding sequence into a 5' half gene and 3' half gene, linked by an artificial intron containing a splicing donor site (SDS) at the 3' end of the 5' half gene and a splicing acceptor site (SAS) at the 5' end of the 3' half gene. See FIG. 28A, which shows a schematic of a GFP based splicing reporters. The Splicing reporter contains a 5'- half GFP and 3-'half GFP, linked by an intron with SDS and SAS indicated.

A mouse Duchenne Muscular Dystrophy gene (Dmd) exon 23 skipping reporter construct was further generated by inserting the mDmd exon 23 and its intronic SAS and SDS sites into the artificial intron sequence of the splicing reporter. As a result, the mDmd exon 23 skipping reporter construct leads to expression of a protein with the exon 23 encoded amino acid sequence between the 5' and 3' half EGFP gene, which is a non-fluorescent protein. By deleting the splicing regulatory site such as exon 23 SDS would cause splicing skipping of exon 23, leading a functional EGFP.

The ME single module system and ME dual module system were designed to delete a 29-bp sequence containing the exon 23/intron 23 SDS region. Similar design was made with

Prime Editors in both the single module PE2 configuration and dual module PE3 configuration. The results are shown in FIG. 28B.

As shown in FIG. 28B, single module ME led to a 40% EGFP expressing cells (indicated as magLow), and adding the second ME module (second module contains only gRNA not magRNA, causing a second nicking) further increased the EGFP producing cells to 70% (indicated as magLow+sgUp).

PE2 produced a similar outcome as the single module ME system, while the efficacy of PE3 (a second nicking was introduced in addition to the first nicking caused by PE2) was higher than PE2 but was lower than the dual module ME.

Example 13

This example (FIG. 29) compared the editing efficacy of a single-ME, a dual-ME with magRNA-gRNA pairing, and a dual-ME with magRNA-magRNA in correcting a deletion mutation when complexed with a long RNA template (Ex111-P14). The ME systems are similar to those illustrated in FIG. 25, except for the last panel experiment, wherein the second gRNA also contains a second matching tag at 3' end.

The results show cells that were either untreated, or electroporated with an mutant EGFP (deletion of one nucleotide at position 50 relative to nicking site) expression vector (EGFP Δ50), or with the mutant EGFP expression vector plus a one-module ME (magLow), or a dual-module ME with magRNA-gRNA pairing (magLow+sgUp119), or a dual-module ME with magRNA-magRNA pairing (magLow +magUp119), all with an RNA template of an extension length of 111 nt. In the magRNA-magRNA pairing, the match tag sequences in the two magRNAs are complementary to two separate sequences in the RNA RT template.

The results showed that the dual-ME system with a second magRNA had the highest editing efficacy, followed by the dual-ME with a magRNA paired with a gRNA, followed by a single-module ME.

Example 14

This example showed a dual ME system editing the endogenous site, HEK3, in HEK293T cells. Importantly, it also showed the different individual contributions of RNA template and magRNA to gene editing efficiency.

A template intended to install two point mutations, 5G>T and 12G>C, was designed to target the HEK3 site as shown in FIG. 30A. The template, T43-P13-(5GT-12GC), contains a

43-nt RT extension sequence and 13-nt primer binding sequence. Both point mutations are transversion mutations, one (5G>T) intended to destroy the PAM motif to avoid repetitive futile editing. Based on the template sequence, magRNA_12M42 was designed with the upper guide as shown in FIG. 30A and a 12-nt complementary to the 5' end of the T43 template. The expression vectors of the template, magRNA, a second nicking gRNA with the guide were shown in FIG. 30A. These plasmids and the nCas9(H840A)-RT fusion plasmids were introduced into HEK293 cells by electroporation. The genomic DNA was extracted from the treated and untreated cells three days later. The DNA fragment containing HEK3 was amplified by PCR and analyzed by Sanger sequencing. FIG. 30B showed that the transversion editing was evident. Moreover, the editing efficiency at 5G>T and 12G>C were similar. The editing efficiency at 5G>T was used as a measurement for quantification (FIG. 30C).

To determine the individual contribution of RNA template and magRNA to editing efficiency, the amount of RNA template and magRNA vectors were varied. As shown in FIG. 30C, keeping the magRNA vector at 250 ng but increasing the RNA template vector from 250 ng to 1000 ng drastically increased editing efficiency. On the other hand, further increasing the magRNA vector from 250 ng to 500 ng had little effect on editing efficacy. Consistently, when the RNA template vector was kept at 250 ng, increasing the magRNA vector to 500 ng did not significantly affect editing efficiency (data not shown).

This example demonstrated that the dual ME system is effective in installing transversion point mutations in the endogenous HEK3 site. Importantly, it demonstrated that the RNA template and gRNA contribute to editing efficiency differently. Optimizing the ratio between the RNA template and magRNA can optimize the genome editing outcome.

As the ratio of 250 ng of magRNA vector, 1000 ng template vector, 250 ng second nicking gRNA vector provided an excellent editing outcome in HEK293 cells, most of our subsequent studies as described below, if not specified, adopted this ratio.

Example 15

This example showed a dual ME system in installing insertion mutations in the endogenous HEK3 site in HEK293 cells. The ME component design was similar to Example 14, except that the templates, T46, T50, and T79, contain a 3-nt, 7-nt, and 36-nt sequence, respectively, for introducing an insertion at +5 position, as shown in FIG. 31A. The 3-nt of CTT is the tri-nucleotide deleted in cystic fibrosis gene CFTR delta 508 allele. The 7-nt of

ACTCAGT is a consensus binding site of transcription factor AP-1. The 36-nt encodes a peptide in the TCR variable region for recognizing an influenza viral epitope.

The efficacy of insertion editing at the HEK3 site was apparent, as shown in FIG. 31B and FIG. 31C. In this experiment, 250 ng magRNA and 1000 ng RNA template vectors were used. Moreover, an extra MS2 was inserted in the gRNA of the second nicking module, which was found to increase editing efficiency for unclear reasons. This example demonstrated that the dual ME is effective in installing insertion at a target endogenous genomic site.

Example 16

This example showed that polymerases other than MMLV-RT can be used in the ME system and that DNA templates can be recruited by magRNA for match editing. A split dual ME system was constructed with components illustrated in FIG. 32A. The nfEGFP with A200G mutation served as a reporter (FIG. 32B). An RNA or DNA template, T29-P14-3GA, was designed to correct the mutation. The magRNA contained an MS2 aptamer at the stem loop region to recruit a polymerase fused with MCP. In addition to the MMLV-RT, R2Bm (a retrotransposon RT), and a variant of Helraiser (a polymerase for DNA transposon Heliton) were fused to MCP. The expression vectors of the trans dual ME were introduced to HEK293 cells by electroporation, and editing efficiency was determined by flow cytometry for the edited fluorescent EGFP expression cells. When the DNA template was used, a single strand DNA oligonucleotide was used instead of an expression vector.

As shown in FIG. 32C, the experiment demonstrated that R2Bm and Helraiser also exhibit significant editing activity, albeit lower than MMLV-RT. Moreover, when Helraiser was used, the DNA template was effective in facilitating the editing, although the efficiency was significantly lower both for MMLV-RT and R2Bm combined with an RNA template.

Example 17

This example showed that an additional effector could be recruited to a dual ME system to enhance editing efficiency. As shown in FIG. 33A, a dual ME system was constructed to target HEK3 site for installing the 2 mutations as described before in FIG. 30. In addition, the second nicking gRNA containing an MS2 aptamer was constructed to recruit a pioneer transcription factor, p65, fused to MCP. The ME systems were introduced into HEK293T cells, and the genome editing efficacy was determined by Sanger sequencing.

As quantified in FIG. 33C, the example demonstrated that recruiting p65 by the second gRNA to the dual ME system significantly increased editing efficiency. Interestingly, without MCP-p65, the presence of MS2 in the second gRNA alone also significantly increased editing efficiency. This phenome was observed in other editing sites and cell lines (data not shown here). The mechanism is unclear, but MS2 might have an effect in stabilizing the second CRISPR complex structure.

Example 18

This example compared the editing efficacy of a split dual ME system and its counterpart split Prime Editing (PE) system on the HEK3 site in HEK293T cells.

As shown in FIG. 34A, a split dual ME system for installing two point mutations at the HEK3 site was constructed with the same template and magRNA as described in FIG. 30. In this dual ME system, the MCP-RT fusion was recruited by the second nicking gRNA, which contains an MS2 aptamer, as shown in FIG. 34C. The counterpart split PE3 was constructed with a pegRNA containing the same guide and RT priming/extension sequences as found in the ME magRNA and RNA template, respectively, as shown in FIG. 34B and 34C.

the BE system was introduced with a 1000ng RNA template vector and 250 ng magRNA vector into the cells. In the same experiment, the counterpart PE system was introduced with a molar equivalent pegRNA vector (equal to 1000ng RNA template vector) into HEK293 cells. The molar ratio of the RNA template in PE system: pegRNA scaffold in PE: RNA template in ME: magRNA scaffold in ME was roughly equal to 1:1:1:0.25. The editing efficiency was determined and illustrated in FIG. 34D.

It was apparent that the split dual ME system exhibited higher editing efficacy than the counterpart PE system at editing the HEK3 site. To rule out the possibility that a high pegRNA vector level may not be preferable for the PE system, the pegRNA level was titrated down to around 250 ng. It was found that the PE system with pegRNA vector at the molar equivalent to 1000ng RNA template vector exhibited higher genome editing efficacy than lower levels of pegRNA (data not shown).

This example demonstrated that the split dual ME system has higher editing efficacy than its counterpart split PE3 system in editing the HEK3 site under their respective favorable conditions.

Example 19

This example was illustrated from FIG. 35 to FIG. 37, which compared a dual ME system and its PE3 counterpart in installing five point mutations at the endogenous HEK3 site in HEK293T cells. In this example, the editing efficiency was determined by Sanger sequencing and Next Generation Sequencing (NGS). Moreover, the unintended insertion formation at the target site was also analyzed with data from NGS.

FIG. 35 illustrates how the dual ME targeting HEK3 was constructed. The magRNA has the same guide and matching tag as those described in Figure 30A. The template, T43-P13-(5G>T-12G>C-18A>C-24T>C-30C>T), however, is different which intends to install five point mutations across the 43-nt RT extension sequence, as illustrated in FIG. 36A. Its counterpart PE3 has a pegRNA containing the same guide as the magRNA and the 3' extension sequence the same as the ME RNA template. The common components of the ME and PE include the second nicking gRNA and the sp-nCas9-RT fusion expression vectors.

The ME and PE systems with the same ratio described in Example 18 (FIG. 34) were introduced into HEK293T cells separately. The second gRNA concentration was also varied in both ME and PE systems (250 ng vs 500 ng). Three days after electroporation, the genomic HEK3 DNA was amplified by PCR and analyzed by Sanger sequencing and by NGS.

As shown in FIG. 36B, the genome editing in installing the 5 point mutations by ME was apparent. Moreover, quantification of the Sanger sequencing results in FIG. 36C showed that the patterns of the five mutations under all the conditions were similar. The editing efficiency with the ME system was higher than the PE system under both second gRNA conditions (250 ng and 500ng). The statistical significance between the ME and PE editing efficiency at 5G>T editing site was shown in FIG. 36D. The same statistical significance was observed in the other 4 mutation sites between the ME and PE systems (data not shown).

NGS of representative samples under these conditions provided a similar but even more precise picture, as shown in FIG. 37A. Among the five point mutations, the editing efficiency decreased as the position's distance to the priming site increased, which is consistent with the editing mechanism. Nevertheless, similar to the conclusion drawn from the Sanger sequencing analysis, NGS results also showed that the ME system exhibited higher editing efficiency than the PE system under both conditions with varying second gRNA amounts.

One unique feature of ME is that the modular ME template is intrinsic "scar-less", while the PE peg template mechanistically tends to incorporate the 3' end gRNA scaffold sequence into the editing end as the reverse transcription continues into the gRNA scaffold. This was

experimentally demonstrated by the NGS data here. The inventor analyzed the insertion of the edited reads at the end of the template sequence and tallied the reads that contain inserts of 3 or more nucleotide bases identical to the 3' end of the gRNA scaffold. As summarized in the table in FIG. 37B, under the condition of a 250 ng second gRNA vector, the dual ME system exhibited only background gRNA scaffold sequence insertion (same as the untreated cells). However, the PE system generated about 350-fold higher the reads containing inserts identical to the gRNA scaffold 3' end. The ratio between these reads to edited reads was about 350-fold higher in the PE system as compared to the ME system (FIG. 37B). When the second gRNA vector increased to 500 ng, the insertion rate of the ME system increased to above the background level, likely due to an increase in DSBs. Nevertheless, the ratio of these reads to edited reads was about 45-fold less than observed with its counterpart PE3 system under the same conditions (FIG. 37B). The striking contrast between ME and PE systems in generating unintended gRNA scaffold sequence insertions at the end of editing site was illustrated in FIG. 37C.

The example demonstrated that the described dual ME system is more efficacious than its counterpart PE3 system in installing multiple point mutations, including both transversion and transition point mutations, at scattered positions simultaneously in HEK3 site. Moreover, the example directly and unequivocally demonstrated that the ME and PE systems have fundamental difference in "scar" formation at the end of editing site. It confirms that the PE system tends to incorporate the 3' end gRNA scaffold sequence into the end of editing site while the ME system does not have the tendency.

Example 20

This example compared a dual ME system and its PE3 counterpart in editing the endogenous HEK3 site by deleting or inserting 3-nt sequences in HEK293T cells. The dual ME system was constructed targeting HEK3 in the same way as described before in FIG. 30, except the template either containing a 3-nt (CTT) insertion or a 3-nt (GCA) deletion, as illustrated in FIG. 38A. Its counterpart PE3 system was similarly constructed with the same template sequence covalently linked to the 3' end of pegRNA.

The dual ME system or its counterpart PE3 was introduced into HEK293T cells and their editing efficacy in installing deletion or insertion was analyzed and compared using Sanger sequencing. Results in FIG. 38B demonstrated that the dual ME system has higher

editing efficacy than its counterpart PE3 system both in installing the 3-nt insertion and 3-nt deletion at the HEK3 site.

Example 21

5 This example showed the efficacy of the ME system in different mammalian cells, K562 cells. The example also compared the editing efficiency of a dual ME system with its counterpart PE3 system in installing nucleotide insertions in episomal EGFP plasmids containing a 1-nt deletion at various positions, *i.e.*, $\Delta 4A$, $\Delta 20C$, and $\Delta 35C$ relative to the EGFP nick site at G202 (FIG. 17C). The dual ME system was constructed as described in FIG. 25
10 (200L + 119U) (the similar experiments were performed in HEK293T cells in FIG. 25). The counterpart PE3 system contains a pegRNA with the same guide and template as the ME. The ME or PE system was electroporated into K562 cells, together with a 1-nt deleted frame-shift EGFP expression plasmid. The gene editing efficiency was quantified by flow cytometry determined by the percentage of cells with fluorescent EGFP expression, as shown in FIG. 39.

15 FIG. 39 demonstrated that the dual ME system effectively installed a one-nucleotide insertion at various deletion sites in K562 cells. Moreover, it showed that the ME system was more efficacious than its PE counterpart in installing these insertions at the $\Delta 20$ and $\Delta 35$ positions. The efficacy of ME and PE was about the same when installing the insertion at $\Delta 4$ position.

Example 22

The example showed that a dual ME system effectively created a 29-nt deletion in an episomal plasmid in K562 cells. In addition, the example compared the editing efficiency of this dual ME system with its counterpart PE system. The ME and PE components and
25 experiments were described in FIG. 28, where the same experiments were performed in HEK293T cells. This example repeated the experiment previously done in HEK293T cells (FIG. 28) in a different mammalian cell type, the K562 cells. The editing efficiency was quantified by the percentage of GFP positive cells (FIG. 40A) and by Sanger sequencing (FIG. 40B). FIG. 40A and FIG. 40B again demonstrated that the dual ME system was more effective
30 than its counterpart PE3 system in generating the 29-nt deletion in K562 cells. Similar observation was made in HEK293T cells in FIG. 28.

Example 23

This example compared a dual ME system and its PE3 counterpart in K562 cells in editing the endogenous HEK3 site by deleting or inserting a 3-nt sequence. Both editing efficacy and the insertion rate caused by RT template extension into the gRNA scaffold were compared. The ME components were the same as described in FIG. 38, where the experiment was performed in HEK293T cells. The edited genomic DNA was subject to Sanger sequencing and Next Generation Sequencing (NGS).

FIG. 41A and 41B showed the efficacy of CTT insertion and GCA deletion is higher with the ME system than the ME system at HEK3 site in K562 cells. Moreover, FIG. 41C shows that the PE system generated two-order of magnitude more gRNA scaffold- templated inserts at the end of the HEK3 editing site than the ME system.

The example demonstrated that the dual ME system effectively installed deletion and insertion in K562 cells. Moreover, it demonstrated that the dual ME system is more efficacious than its counterpart PE3 system in installing the 3-nt deletion and insertion at this HEK3 site. Importantly, the example again directly and unequivocally demonstrated, in a different mammalian cell type, that the ME and PE systems have fundamental differences in “scar” formation at the end of the editing site. The tendency of PE system to incorporate the 3’ end gRNA scaffold sequence into the editing end is about two-order of magnitude higher than the ME system.

Example 24

The example showed a dual ME system is efficacious in installing a mutation at a therapeutically relevant genomic site, the endogenous HBB (hemoglobin beta) gene site near the sickle cell anemia E6V mutation, in K562 cells. FIG. 42A shows the HBB sequence near the E6V (GAG>GTG) mutation site, with annotations of magRNA guide, second nick guide, nicking sites, as well as the intended +5G>T mutation encoded in the RNA template. The A base immediately 5’ to +5G was mutated in sickle cell anemia E6V.

The dual ME system was electroporated into K562 cells and the editing was analyzed using Sanger sequencing. FIG. 42B demonstrated that the dual ME can install a +5G>T transversion base change at the endogenous therapeutic site of HBB at high efficiency in K562 cells.

Example 25

This example tested whether ME can use CRISPR complex other than the spCas9 system. The example showed that the Cas9 ortholog, saCas9, is effective when used in a dual ME system. Moreover, the example showed that the dual saCas9 ME system is more effective
5 than its saPE3 counterpart at editing an EGFP mutation.

A dual ME was constructed with the Cas9 ortholog, nickase saCas9 (N580A)-RT fusion. Accordingly, the saCas9 gRNA scaffold was used to construct magRNA and the second nicking gRNA. The gene to be edited is an EGFP with a one-nucleotide deletion in the EGFP coding position 100, Δ 100G. FIG. 43A illustrates the targeting sequence, saCas9 PAMs,
10 magRNA guide, and the second nicking gRNA guide. The template was designed to install the missing G at position 100 and a separate silence mutation. For the PE3 system, the pegRNA used the same guides as magRNA, and the 3' RT extension and Primer binding sequences are identical to the RNA template sequence of ME.

The ME and PE systems were electroporated into HEK293, together with the EGFP
15 (Δ 100G) expression vector. Three days later, the cells were subject to flow cytometry assay to quantify the percentage of cells expressing fluorescent EGFP (FIG. 43B-C). The EGFP plasmid DNA surrounding the deletion region was amplified by PCR and analyzed by Sanger sequencing (FIG. 43D). FIG. 43B showed the apparent editing by the saCas9 dual ME system in correcting the deletion mutation. FIG. 43C and FIG. 43D demonstrated that the dual saCas9
20 ME is more efficacious than its saPE3 counterpart.

The foregoing examples and description of the preferred embodiments should be taken as illustrating, rather than as limiting the present disclosure as defined by the claims. As will be readily appreciated, numerous variations and combinations of the features set forth above can be utilized without departing from the present disclosure as set forth in the claims. Such
25 variations are not regarded as a departure from the scope of the disclosure, and all such variations are intended to be included within the scope of the following claims. All references cited herein are incorporated by reference in their entireties.

CLAIMS

WHAT IS CLAIMED IS:

1. A gene editing complex for editing a target site in a target DNA molecule, comprising
 - (A) an RNA-guided nickase;
 - (B) a reverse transcriptase;
 - (C) an RNA template molecule comprising
 - (1) a template segment comprising a template sequence complementary to a desired DNA sequence to be introduced into the target site, and
 - (2) a priming segment complementary to the 3' end of the nicking strand of the target DNA molecule; and
 - (D) a matching gRNA (magRNA) molecule comprising
 - (1) a guide or spacer sequence complementary to a target sequence on a target strand of the target DNA molecule,
 - (2) an RNA scaffold capable of binding to the RNA-guided nickase, and
 - (3) an anchoring tag sequence complementary to a segment in the RNA template.
2. The complex of claim 1, wherein the reverse transcriptase is covalently linked to the RNA-guided nickase.
3. The complex of claim 1, wherein the magRNA further comprises a protein-binding motif capable of binding to an RNA-interacting protein, and the reverse transcriptase is linked to the RNA-interacting protein.
4. The complex of any one of claims 1-3, wherein the anchoring tag in magRNA is not a polyN, wherein N is a repeating nucleotide A, C, U, or G.
5. The complex of any one of claims 1-4, wherein the RNA template does not comprise additional sequence other than the said priming segment and template segment.

6. The complex of claim 1, wherein the RNA-guided nickase is a nickase variant of a Cas protein or a nickase variant of transposon-encoded IscB, IsrB, or TnpB family endonuclease protein.
7. The complex of claim 6, wherein the Cas protein is selected from Cas3, Cas4, Cas5, Cas5e (or CasD), Cas6, Cas6e, Cas6f, Cas7, Cas8a1, Cas8a2, Cas8b, Cas8c, Cas9, Cas10, Cas10d, Cpf1 (Cas12a), C2c1 (Cas12b), C2c3 (Cas12c), CasY (Cas12d), CasX (Cas12e), Cas14 (Cas12f), CasPhi (Cas12j), Cas13a, Cas13b, Cas13c, Cas13d, Cas13x, CasF, CasG, CasH, Csy1, Csy2, Csy3, Cse1 (or CasA), Cse2 (or CasB), Cse3 (or CasE), Cse4 (or CasC), Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csz1, Csx15, Csf1, Csf2, Csf3, Csf4, Cu1966, and the orthologues thereof.
8. The complex of claim 6, wherein the Cas protein is Cas9 and the orthologues thereof.
9. The complex of claim 6, wherein the nickase variant of Cas protein is *Streptococcus pyogenes* nCas9(H840A) or *Staphylococcus aureus* nCas9 (N580A).
10. The complex of any one of the preceding claims, wherein the reverse transcriptase is a naturally-occurring reverse transcriptase from a retrovirus or a retrotransposon, or a variant thereof.
11. The complex of claim 10, wherein the reverse transcriptase is selected from the group consisting of Moloney Murine Leukemia Virus (M-MLV), Human Immunodeficiency Virus (HIV) reverse transcriptase, Avian Sarcoma-Leukosis Virus (ASLV) reverse transcriptase, Rous Sarcoma Virus (RSV) reverse transcriptase, Avian Myeloblastosis Virus (AMV) reverse transcriptase, Avian Erythroblastosis Virus (AEV) Helper Virus MCAV reverse transcriptase, Avian Myelocytomatosis Virus MC29 Helper Virus MCAV reverse transcriptase, Avian Reticuloendotheliosis Virus (REV-T) Helper Virus REV-A reverse transcriptase, Avian Sarcoma Virus UR2 Helper Virus UR2AV reverse transcriptase, Avian Sarcoma Virus Y73 Helper Virus YAV reverse transcriptase, Rous Associated Virus (RAV) reverse transcriptase, and Myeloblastosis Associated Virus (MAV) reverse transcriptase, Line 1 ORF2, R2Bm, and R2O1.

12. The complex of claim 10 or claim 11, wherein the reverse transcriptase is an MMLV-RT or a variant thereof.
13. The complex of any one of claims 1-12, wherein the anchoring tag sequence is at the 3' or 5' end of the magRNA.
14. The complex of claim 13, wherein the anchoring tag sequence is covalently linked to the magRNA via a polynucleotide linker.
15. The complex of any one of the preceding claims, wherein the anchoring tag sequence in the magRNA is about 6-24 nt. in length.
16. A system for editing a target site in a target DNA molecule, comprising
 - (I) a first gene editing complex of any one of claims 1-15, and
 - (II) a second gene editing complex.
17. The system of claim 16, wherein the second gene editing complex comprises
 - (A) a second RNA-guided nickase, and
 - (B) a gRNA molecule comprising a second guide or spacer sequence and a second RNA scaffold capable of binding to the second RNA-guided nickase
18. The system of claim 16, wherein the second gene editing complex comprises
 - (A) a second RNA-guided nickase; and
 - (B) a second magRNA molecule comprising
 - (1) a second guide or spacer sequence,
 - (2) a second RNA scaffold capable of binding to the second RNA-guided nickase, and
 - (3) a second anchoring tag sequence.
19. The system of claim 17 and 18, wherein the second gene editing complex comprises
 - (C) a second reverse transcriptase.

20. The system of any one of claims 16 to 19, wherein the guide sequence of the magRNA in the first gene editing complex and the second guide sequence of the gRNA or second magRNA in the second gene editing complex are complementary to two target sequences of the target DNA molecule, respectively.
21. The system of any one of claims 18 to 20, wherein the second anchoring tag sequence of the second magRNA is complementary to a second segment within the RNA template.
22. The system of any one of claims 16 to 21, wherein the 5' end of the RNA template comprises a sequence identical to the 3' end of the strand of the target DNA molecule that is nicked by the second RNA-guided nickase.
23. The system of any one of claims 19 to 20, wherein the second gene editing complex further comprises (D) a second RNA template.
24. The system of any one of claims 19 to 23, wherein the second reverse transcriptase is linked to the second RNA-guided nickase.
25. The system of any one of claims 19 to 23, wherein the second magRNA molecule or the gRNA molecule comprises a second protein-binding motif capable of binding to a second RNA-interacting protein, and the second reverse transcriptase is linked to the second RNA-interacting protein.
26. A method of modifying a target DNA molecule in a cell, comprising contacting the target DNA molecule with the gene editing complex of any one of claims 1-15 or the system of any one of claims 16-25.
27. The method of claim 26, wherein said modifying leads to a point mutation to the target DNA molecule, an insertion to the target DNA molecule, a deletion to the target DNA molecule, or a combination thereof.
28. The method of any one of claims 26 to 27, wherein the cell is selected from the group consisting of: an archaeal cell, a bacterial cell, a eukaryotic cell, a eukaryotic single-cell organism, a somatic cell, a germ cell, a stem cell, a plant cell, an algal cell, an animal cell, in

invertebrate cell, a vertebrate cell, a fish cell, a frog cell, a bird cell, a mammalian cell, a pig cell, a cow cell, a goat cell, a sheep cell, a rodent cell, a rat cell, a mouse cell, a non-human primate cell, and a human cell.

29. The method of any one of claims 26 to 28, wherein the cell is in or derived from a human or non-human subject.

30. A genetically engineered cell obtained according to the method of any one of claims 26 to 29 or a progeny thereof.

31. The cell of claim 30 wherein the cell is selected from the group consisting of an archaeal cell, a bacterial cell, a eukaryotic cell, a eukaryotic single-cell organism, a somatic cell, a germ cell, a stem cell, a plant cell, an algal cell, an animal cell, in invertebrate cell, a vertebrate cell, a fish cell, a frog cell, a bird cell, a mammalian cell, a pig cell, a cow cell, a goat cell, a sheep cell, a rodent cell, a rat cell, a mouse cell, a non-human primate cell, and a human cell.

32. The cell of claim 30 to 31 wherein the cell is selected from the group consisting of pluripotent stem cells (PSCs), adult stem cells (ASCs), fibroblasts, chondrocytes, keratinocytes, hepatocytes, pancreatic islet cells, and immune cells, including T cells, dendritic cells (DCs), natural killer (NK) cells, and macrophages derived from a human or non-human subject.

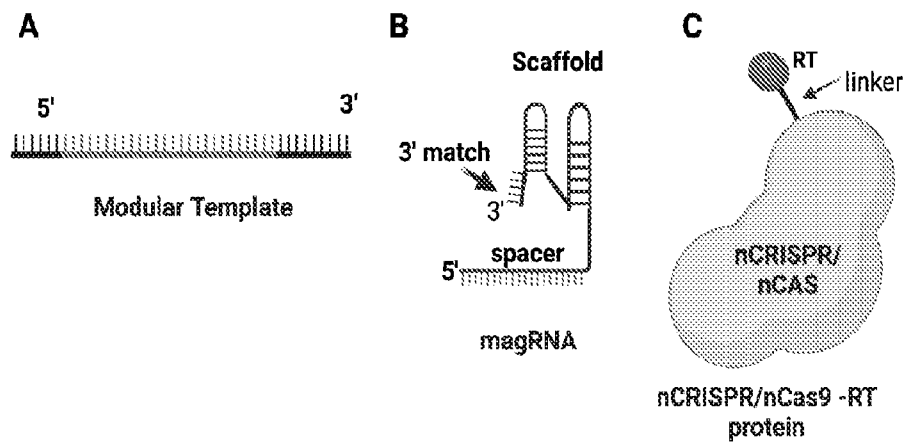
33. A magRNA as recited in claim 1, 3, 4, 13, 14, or 15.

34. An RNA complex comprising the magRNA of claim 33 and the magRNA- anchored RNA template recited in claim 1, 5, 21, 22, or 23.

35. A nucleic acid encoding one or two of:
(i) the magRNA molecule of claim 33, and
(ii) the RNA complex of claim 34.

36. A vector comprising the nucleic acid of claim 35.

37. A kit comprising (i) a packaging material, and (ii) one, two, or more of:
the gene editing complex of any one of claims 1 to 15,
the system of any one of claims 16 to 25,
the cell of claim 30, 31 or 32,
the magRNA molecule of claim 33,
the RNA complex of claim 34,
the nucleic acid of claim 35, and
the vector of claim 36.
38. A pharmaceutical composition comprising (i) a pharmaceutically acceptable carrier,
and (ii) one, two, or more of:
the gene editing complex of any one of claims 1 to 15,
the system of any one of claims 16 to 25,
the cell of claim 30, 31, or 32
the magRNA molecule of claim 33,
the RNA complex of claim 34,
the nucleic acid of claim 35, and
the vector of claim 36.



FIGS. 1A, 1B, and 1C

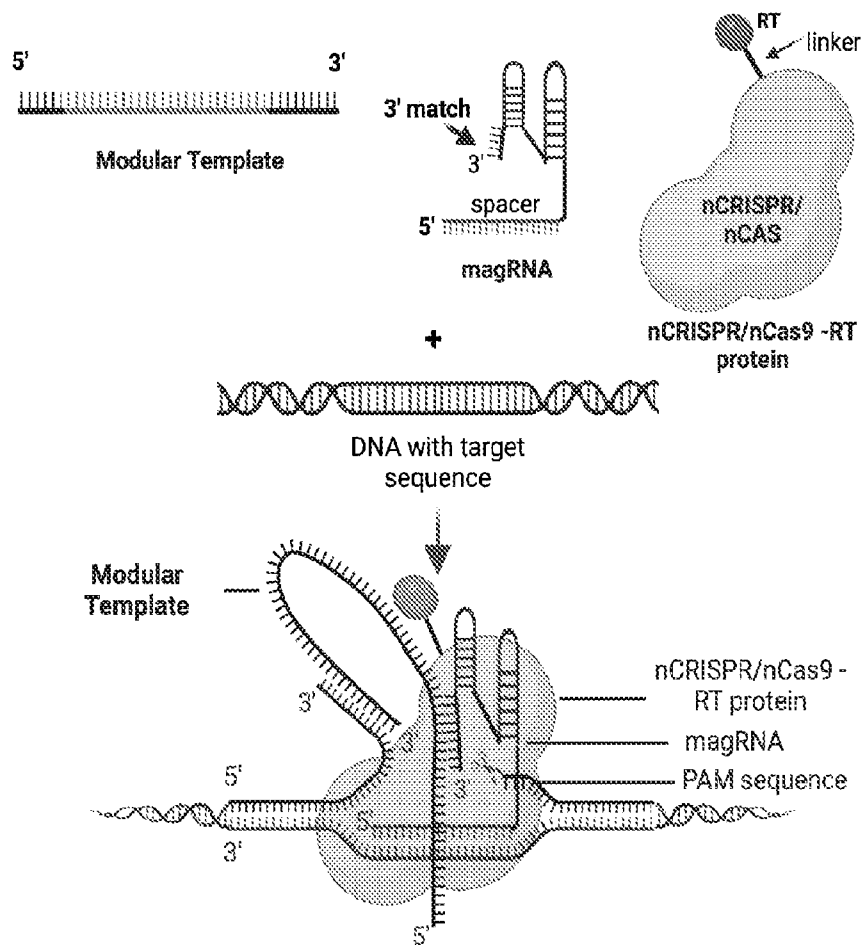


FIG. 2

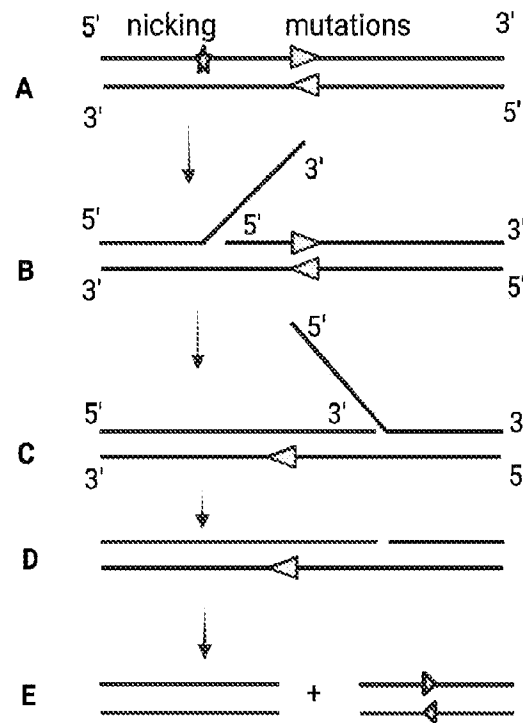


FIG. 3

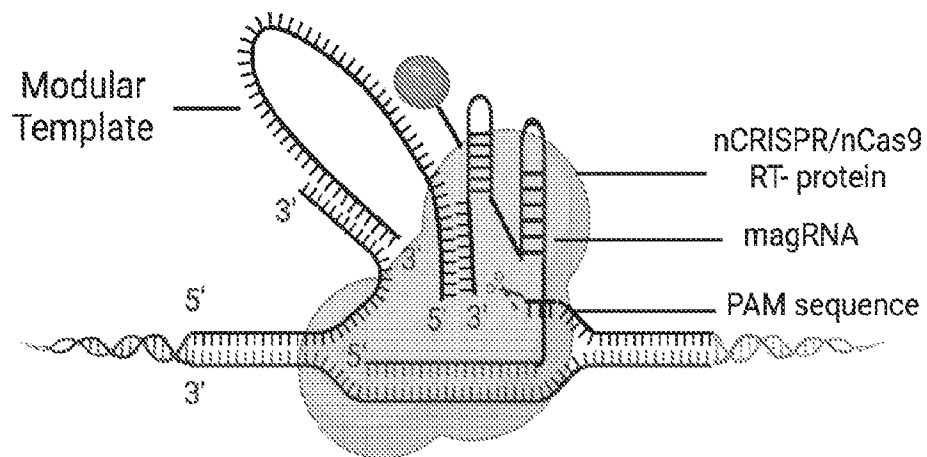


FIG. 4

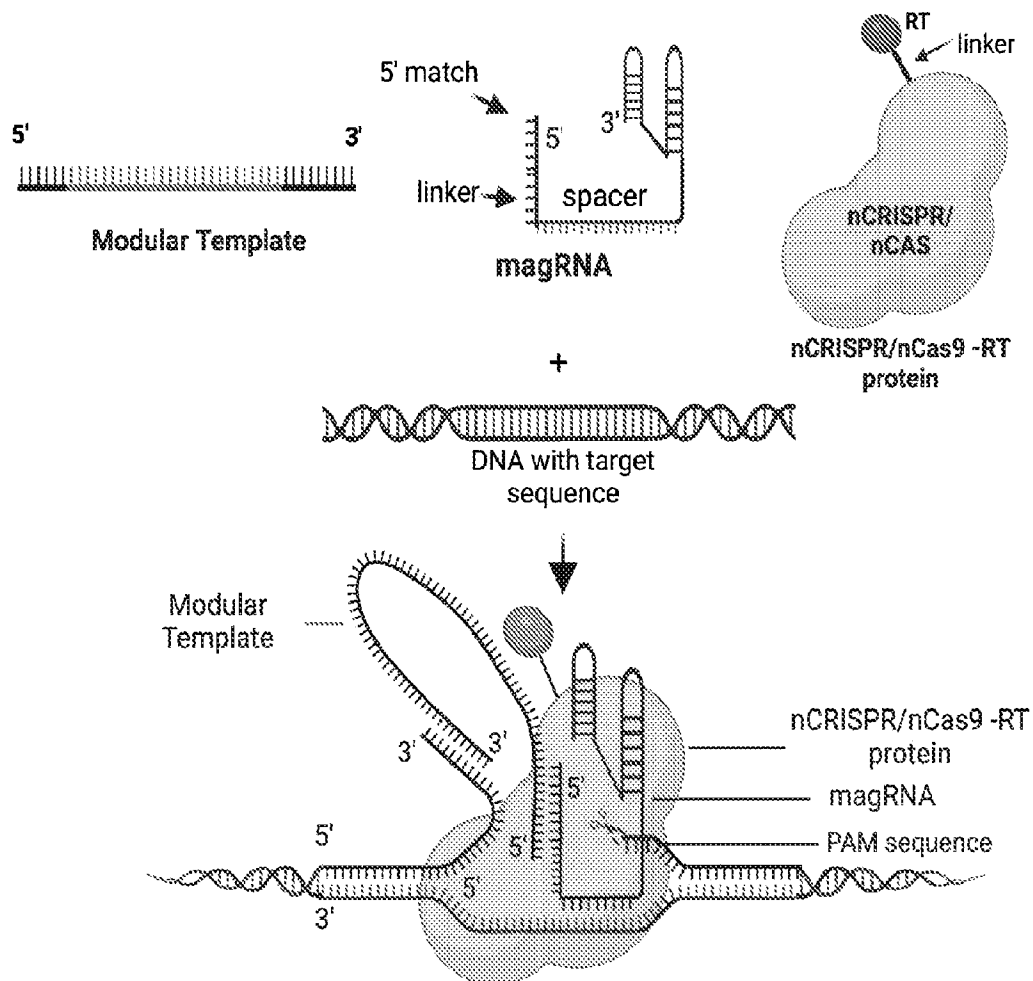
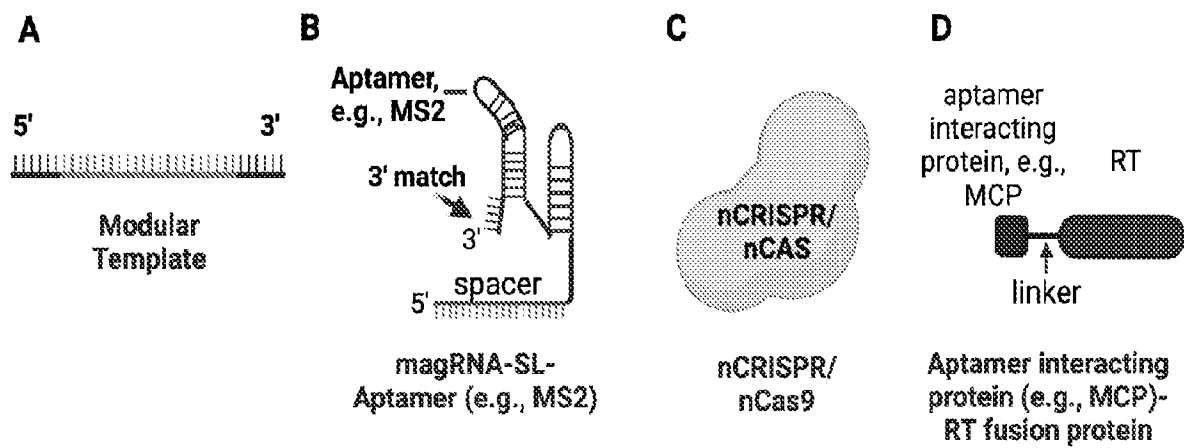
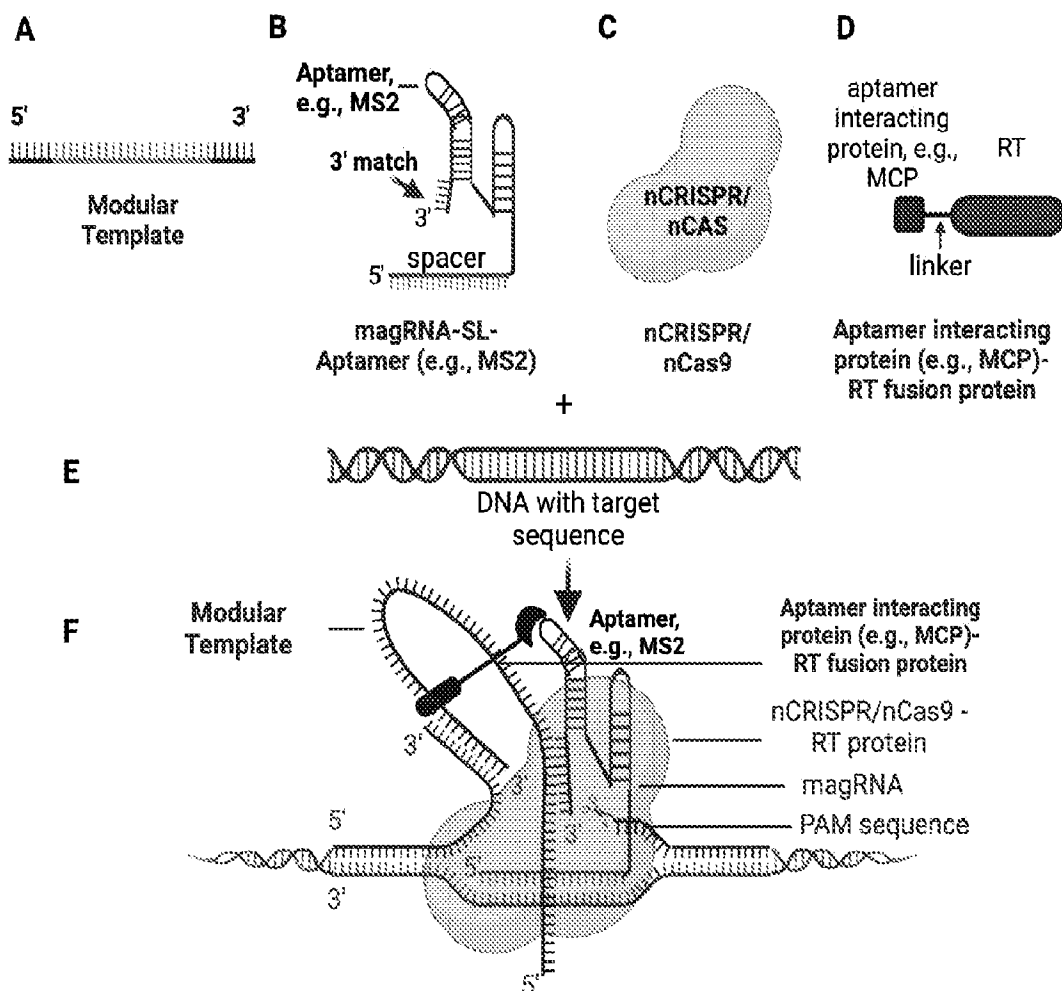


FIG. 5



FIGs. 6A, 6B, 6C, and 6D



FIGs. 7A, 7B, 7C, and 7D

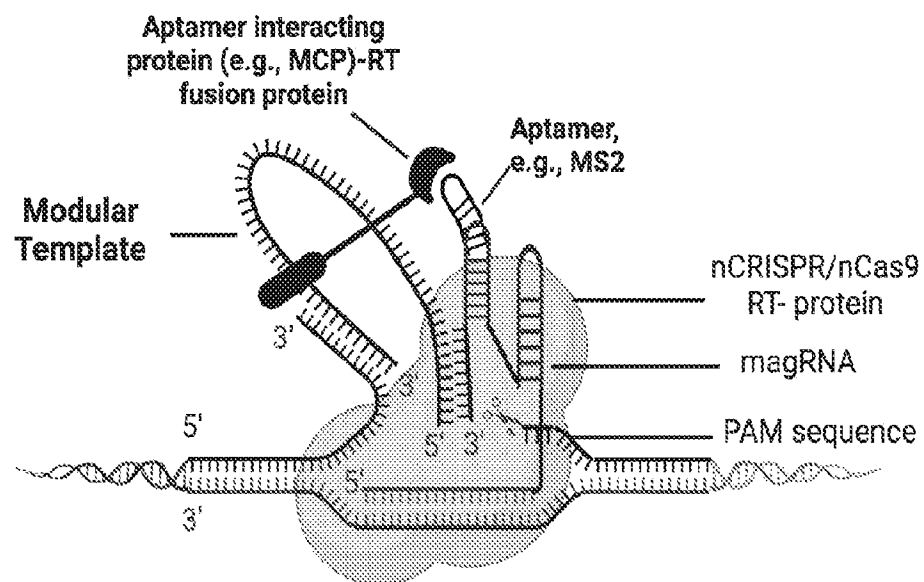


FIG. 8

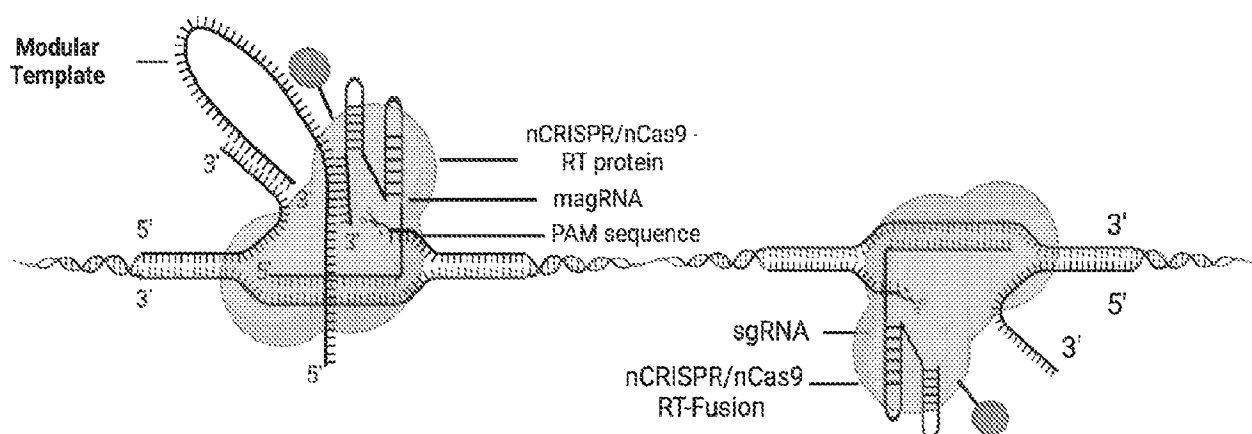


FIG. 9A

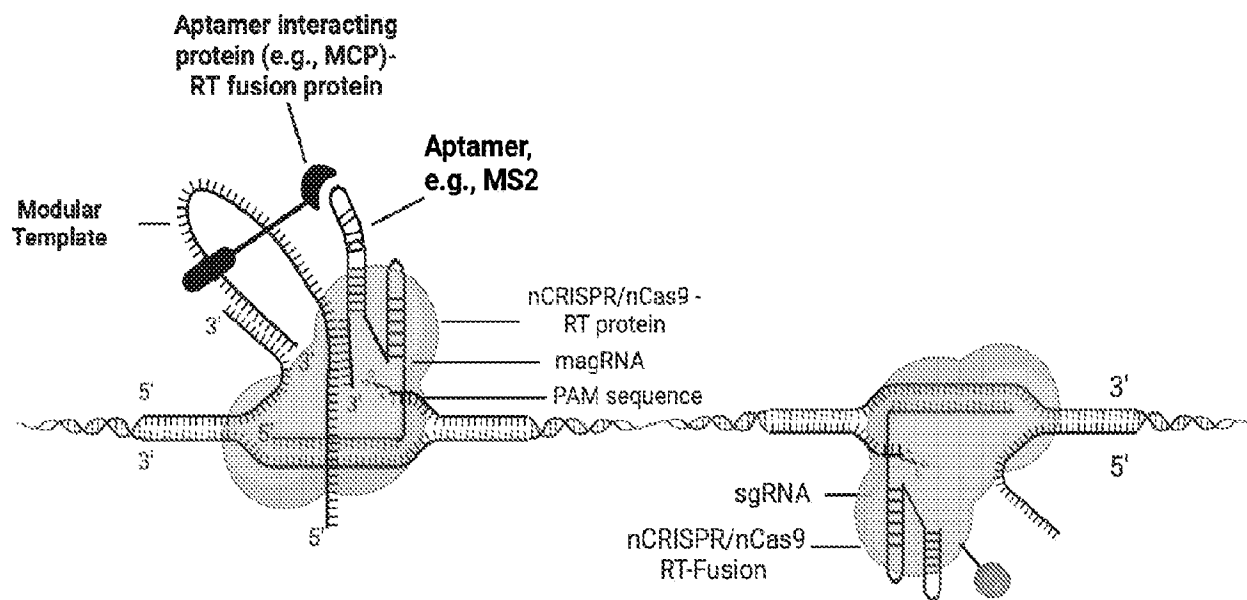


FIG. 9B

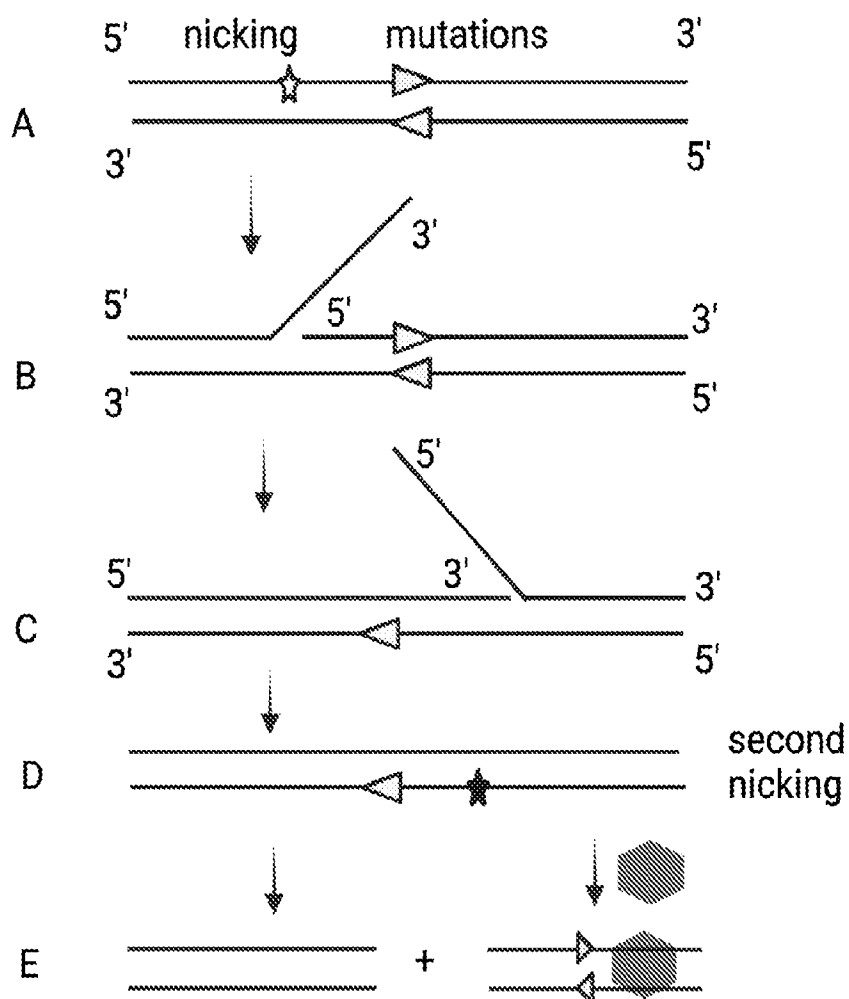


FIG. 10

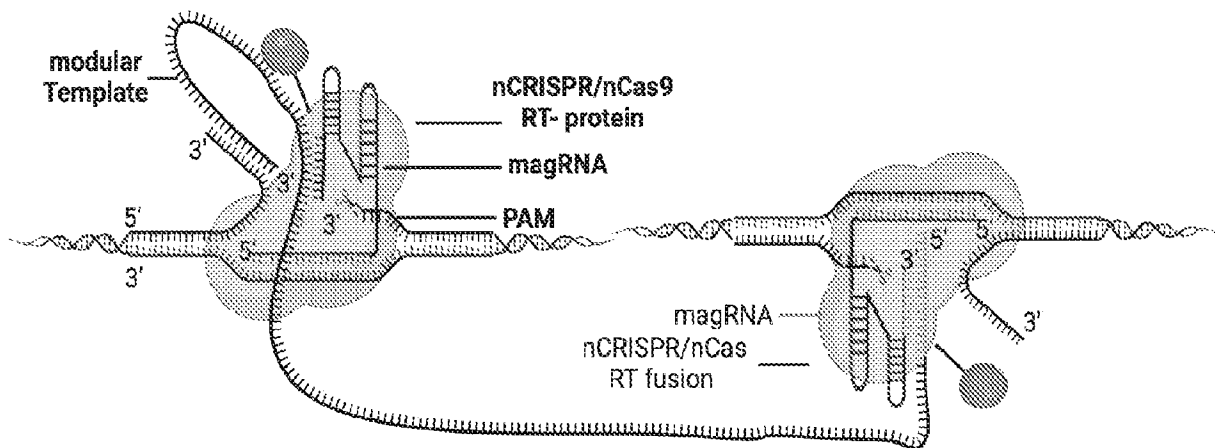


FIG. 11A

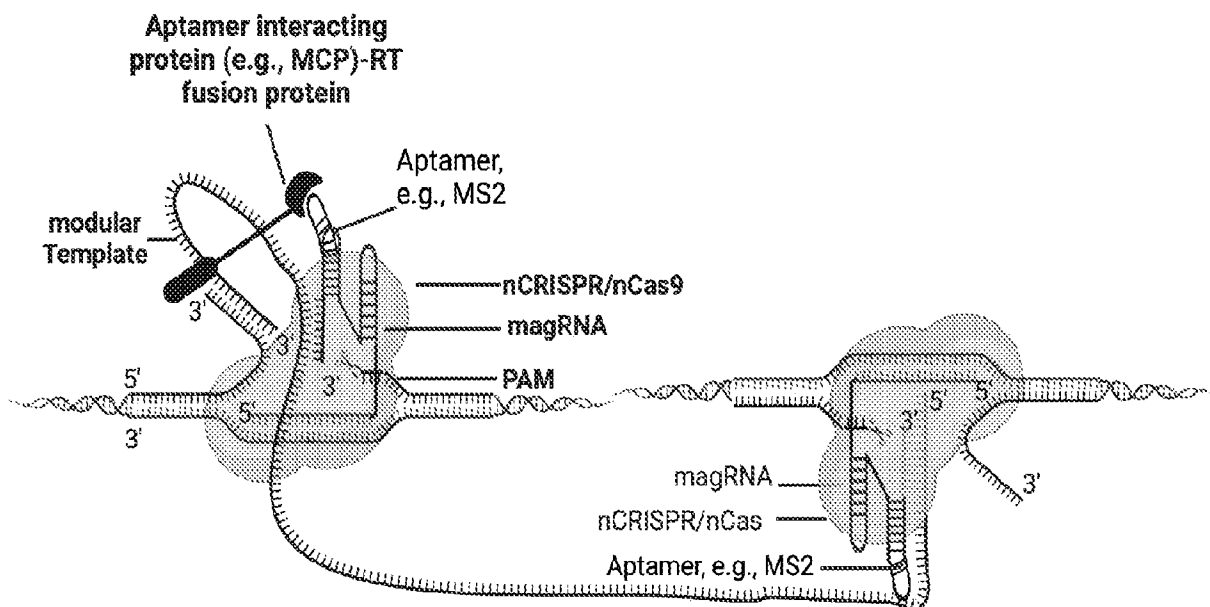


FIG. 11B

Dual ME- mag/mag second priming

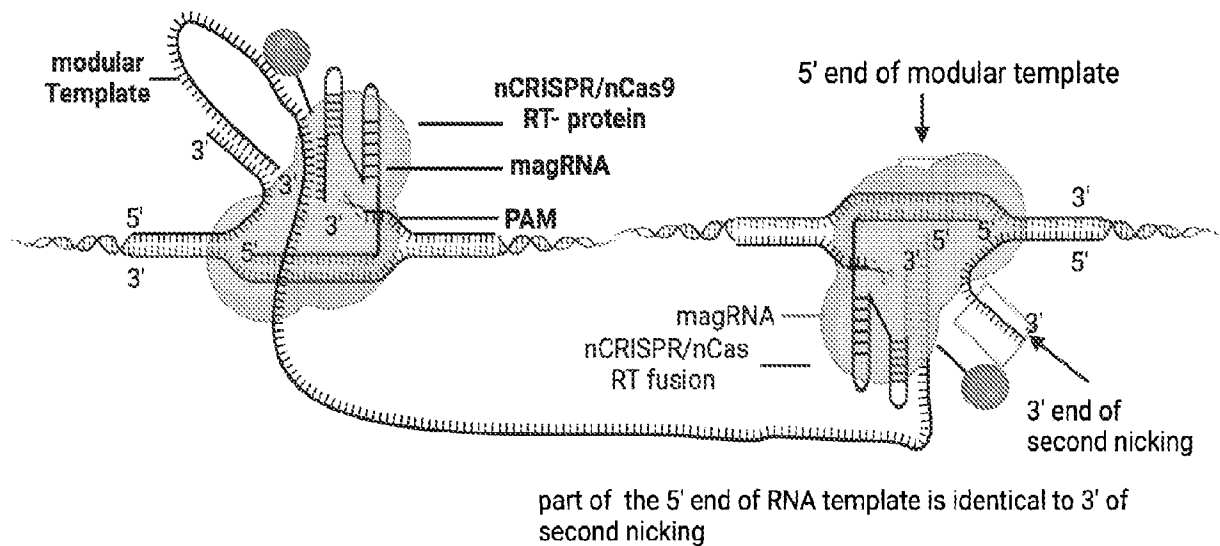


FIG. 12A

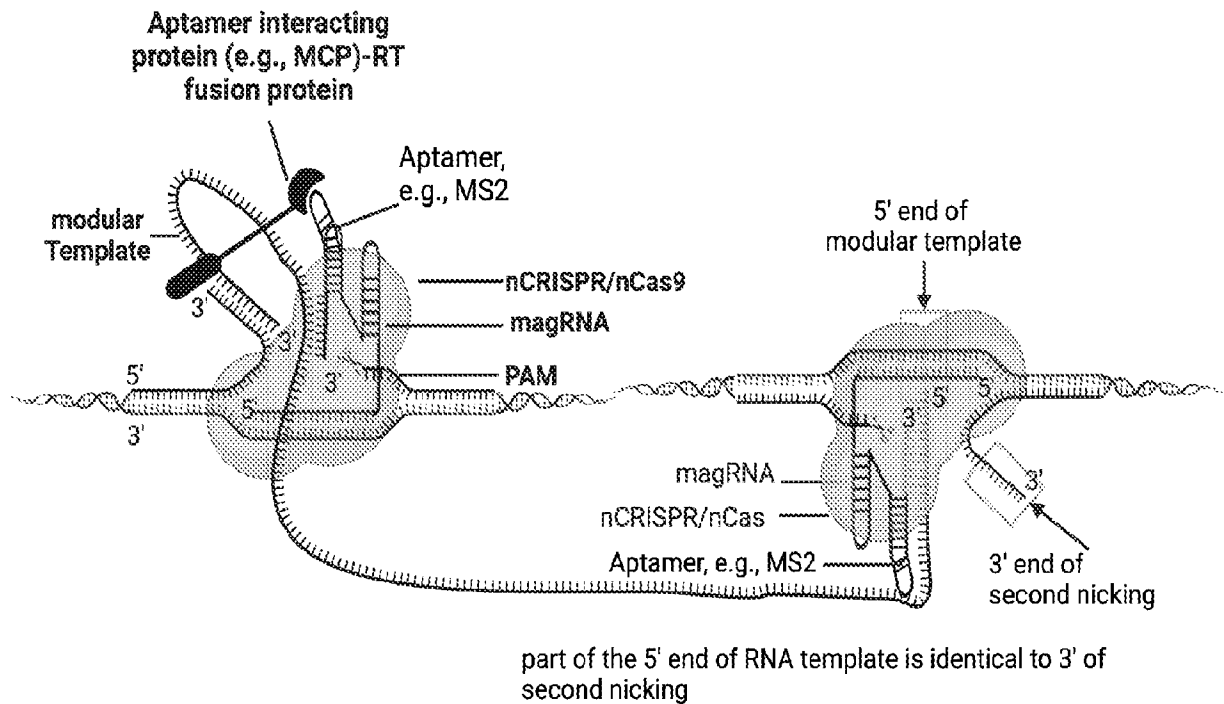


FIG. 12B

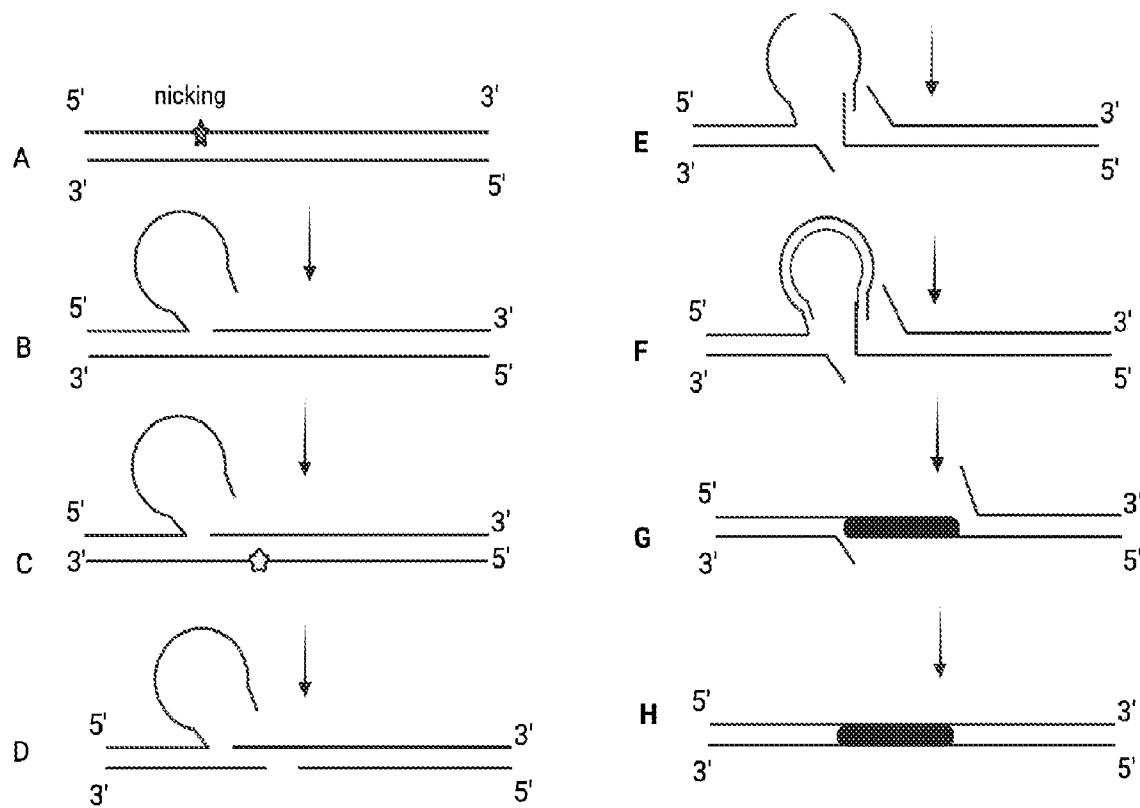


FIG. 13

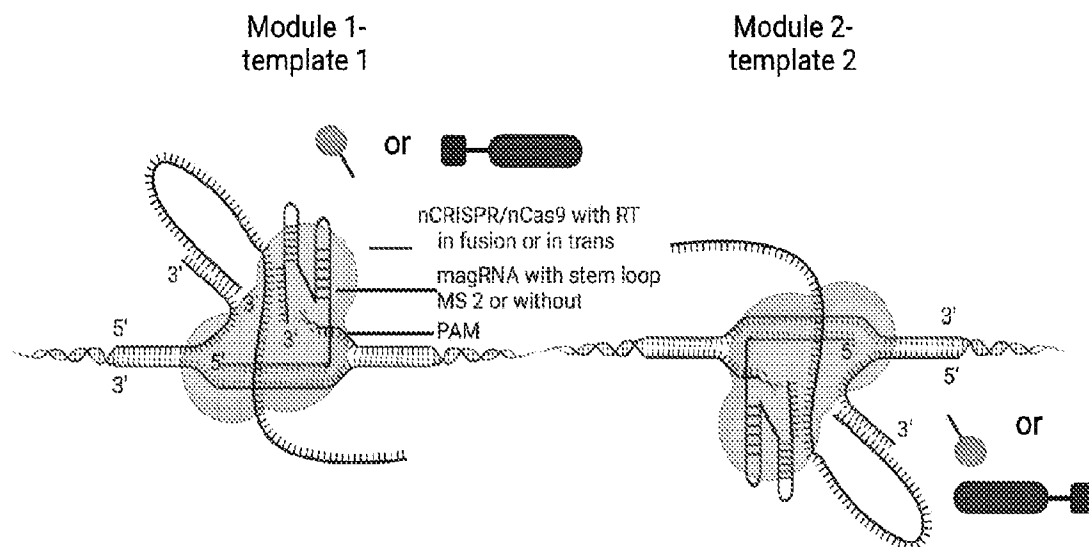
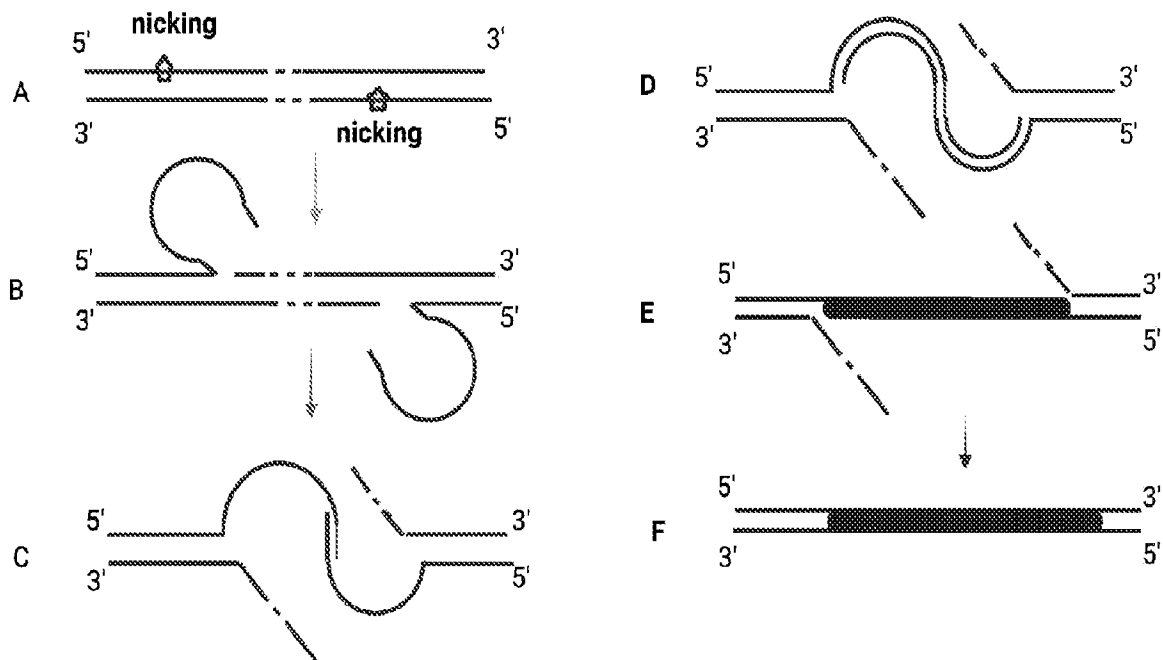
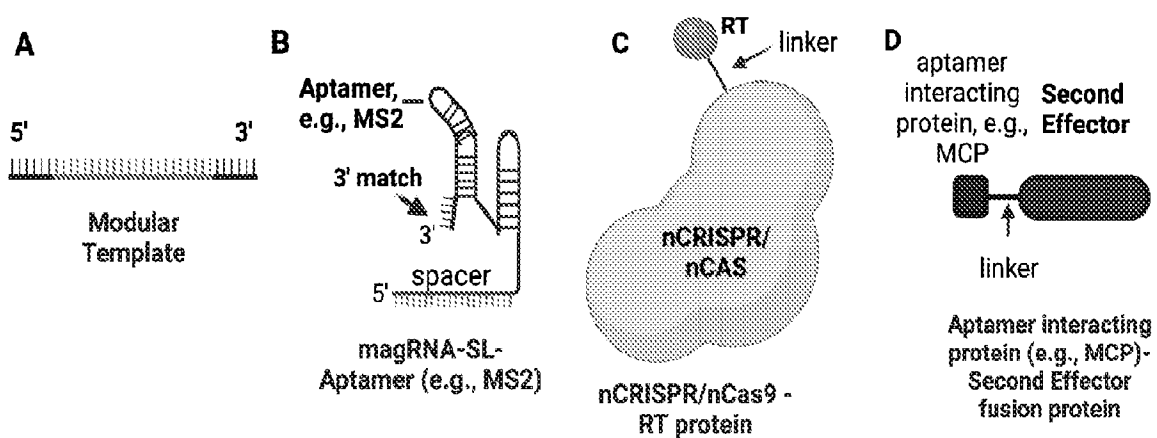


FIG. 14

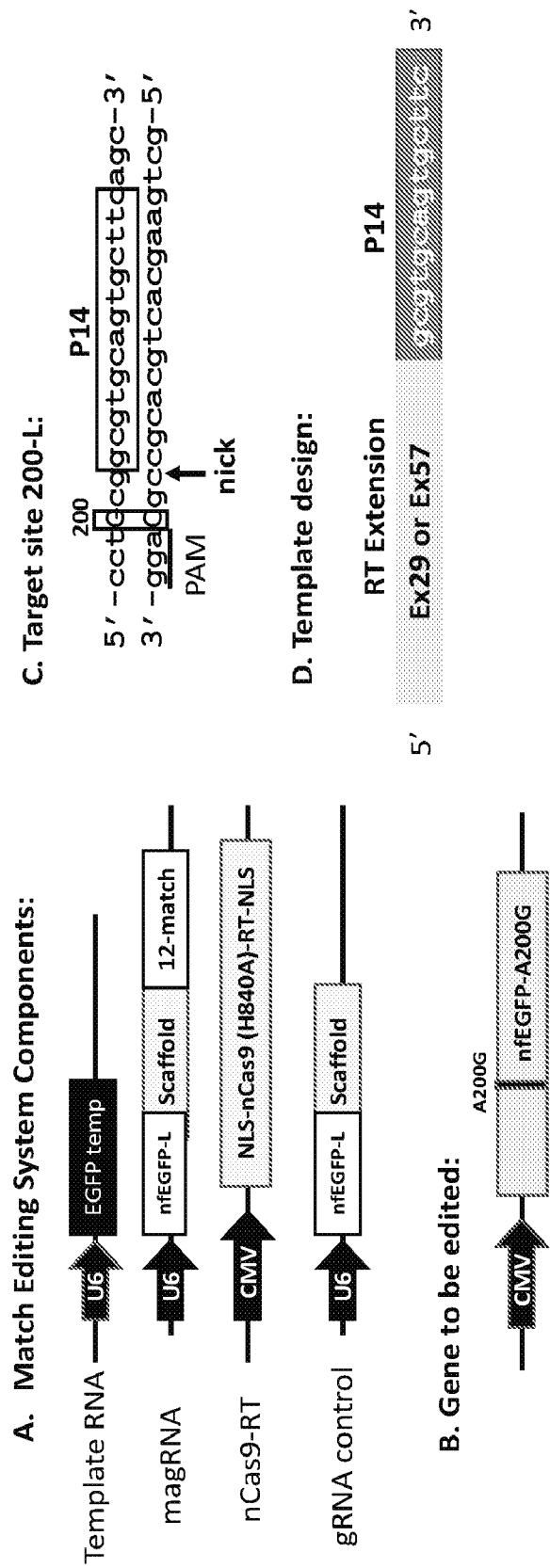


FIGs. 15A, 15B, 15C, 15D, 15E, and 15F

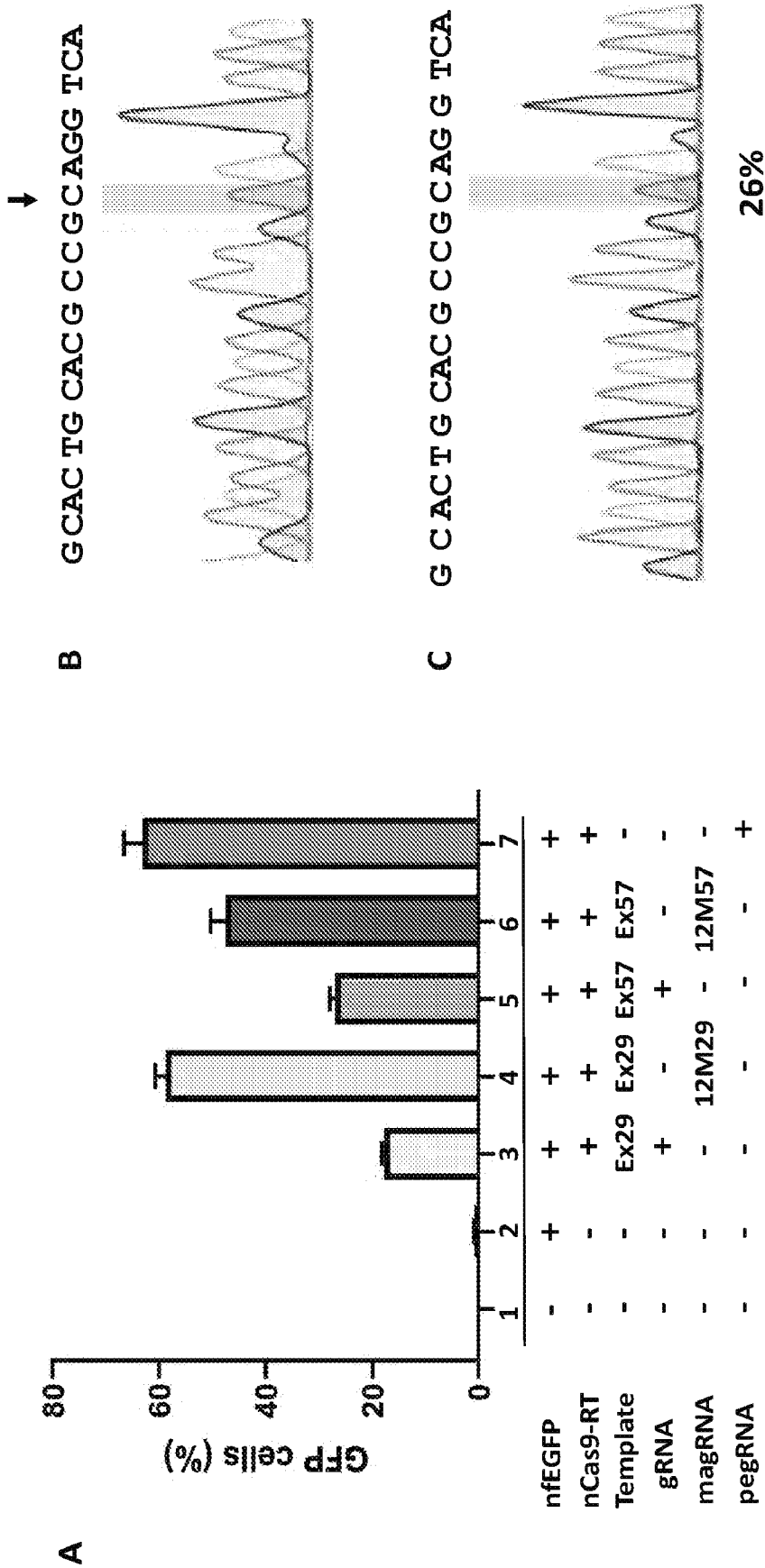


FIGs. 16A, 16B, 16C, and 16D

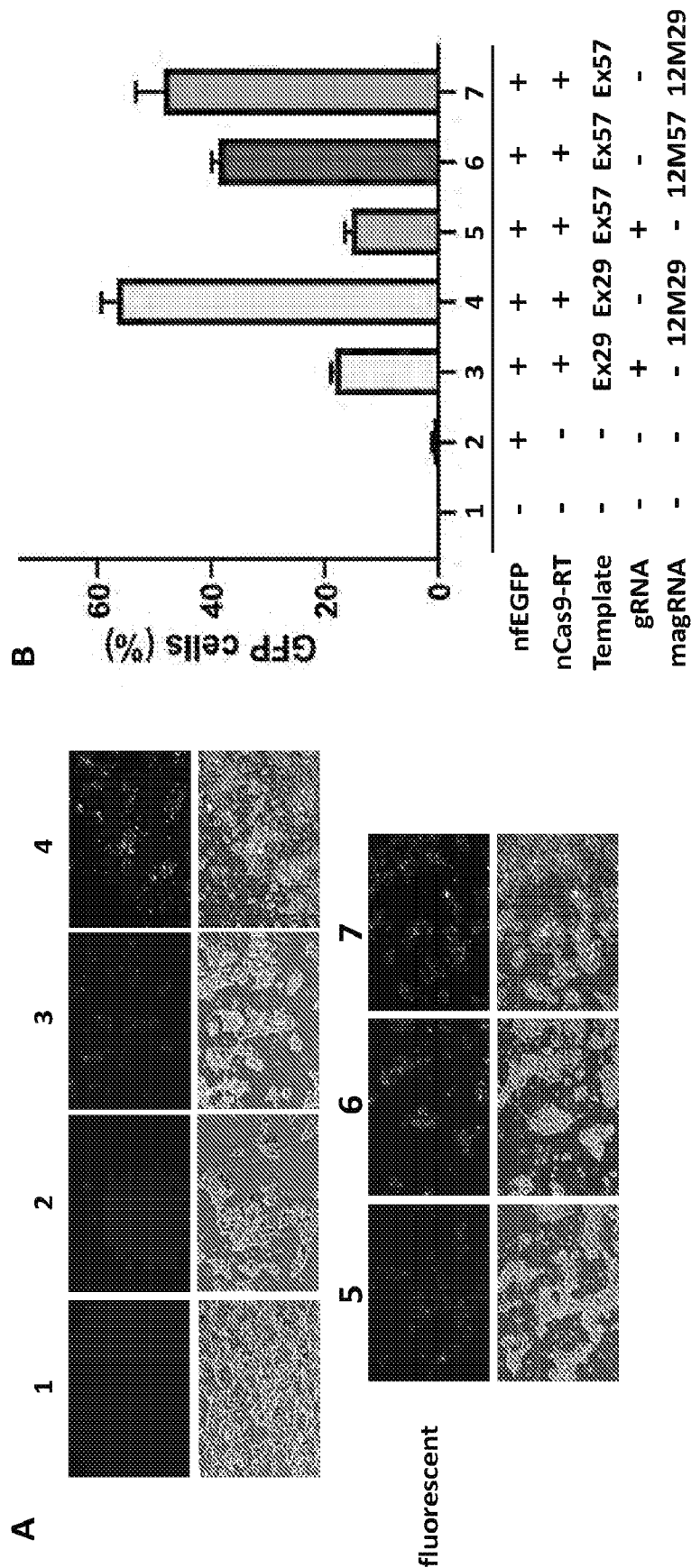
FIGs 1-16 were created with BioRender.com



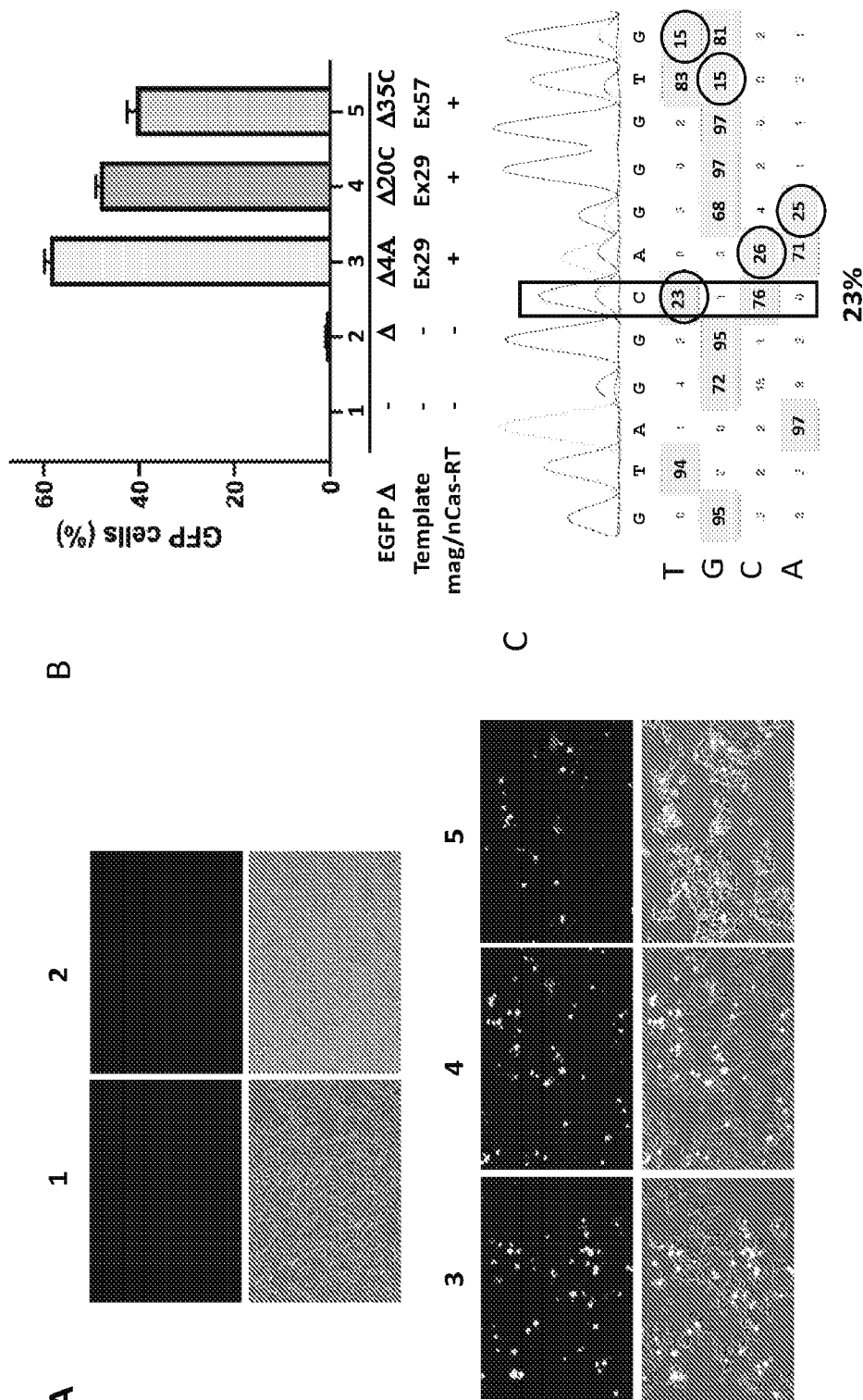
FIGs. 17A, 17B, 17C, and 17D



FIGs. 18A, 18B, and 18C



FIGs. 19A and 19B



FIGs. 20A, 20B and 20C

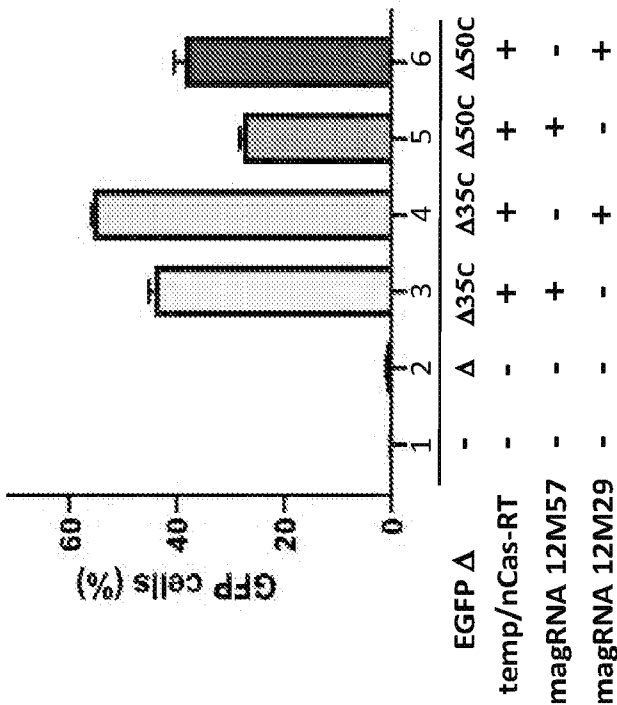


FIG. 21

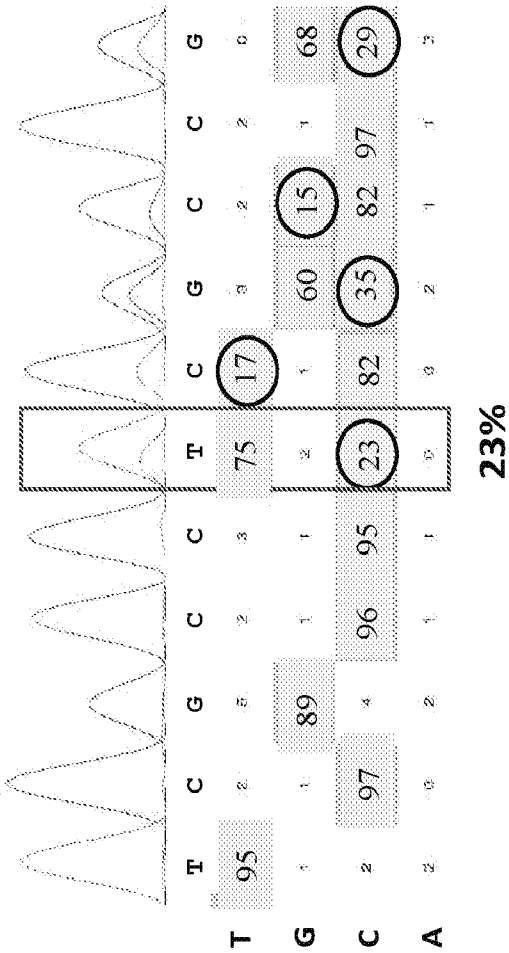
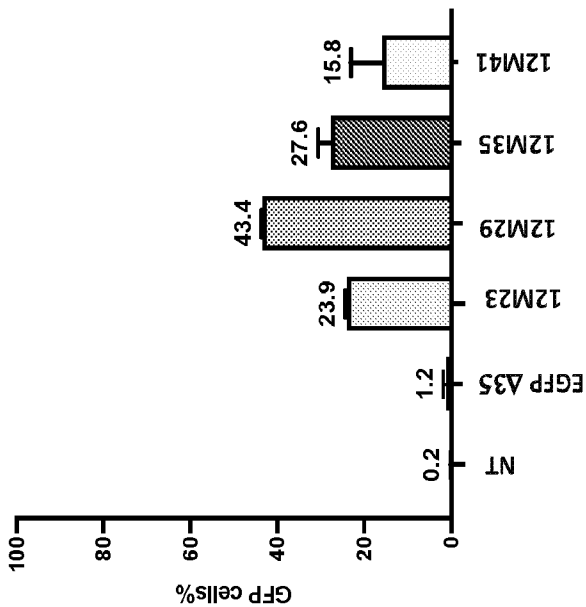
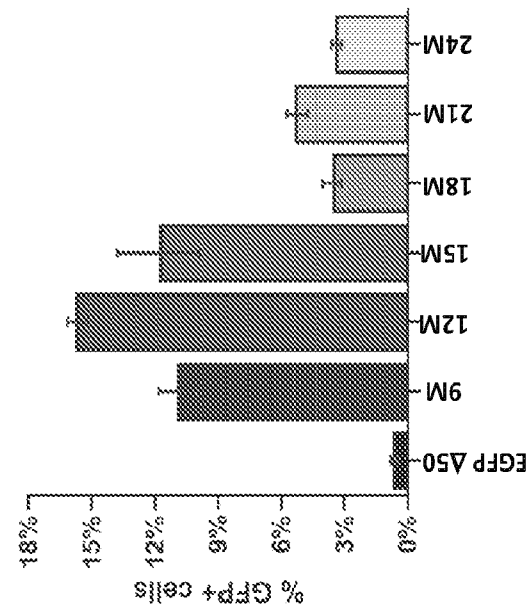


FIG. 22



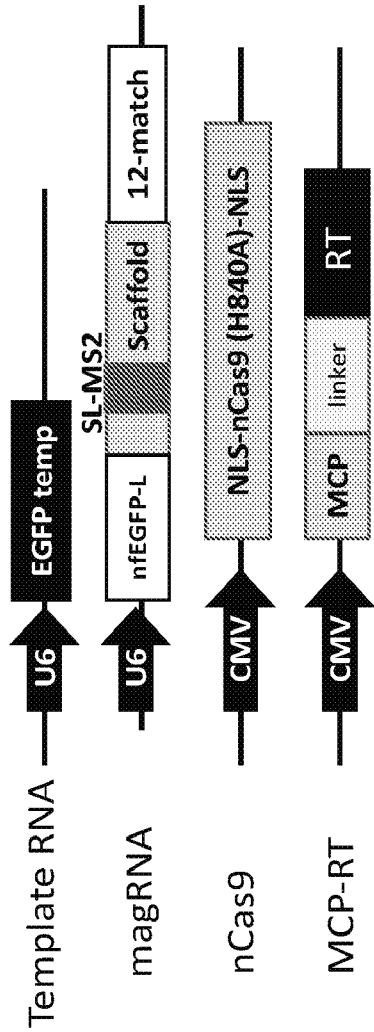
B



A

FIGs. 23A and 23 B

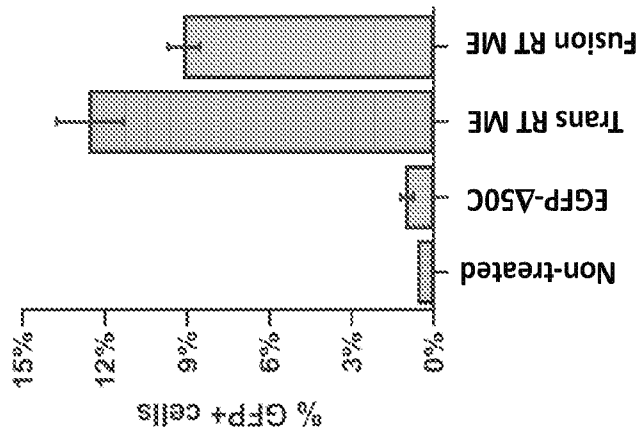
A. Trans RT Match Editing System Components:



B. Gene to be edited:



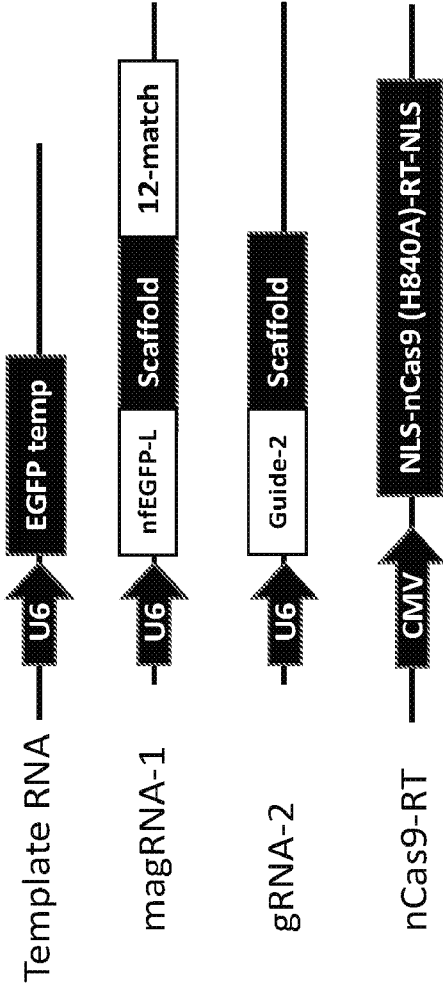
C.



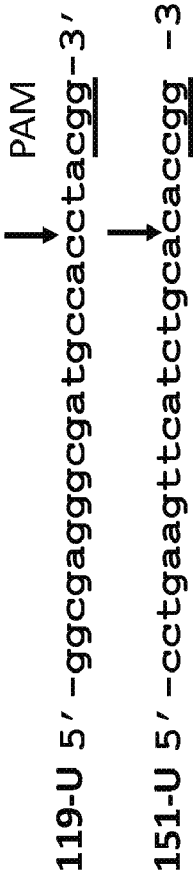
FIGs. 24A, 24B, and 24C

C.

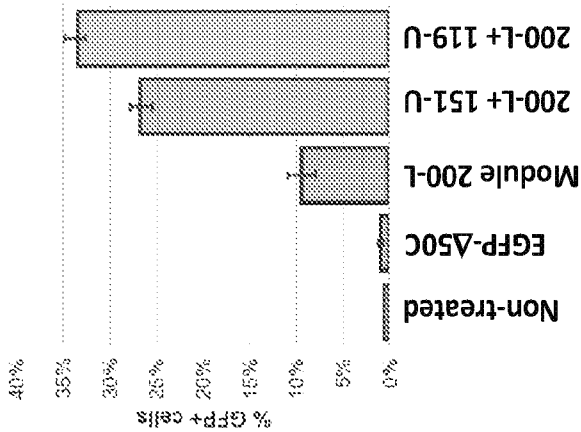
A. Dual ME magRNA-gRNA pairing system



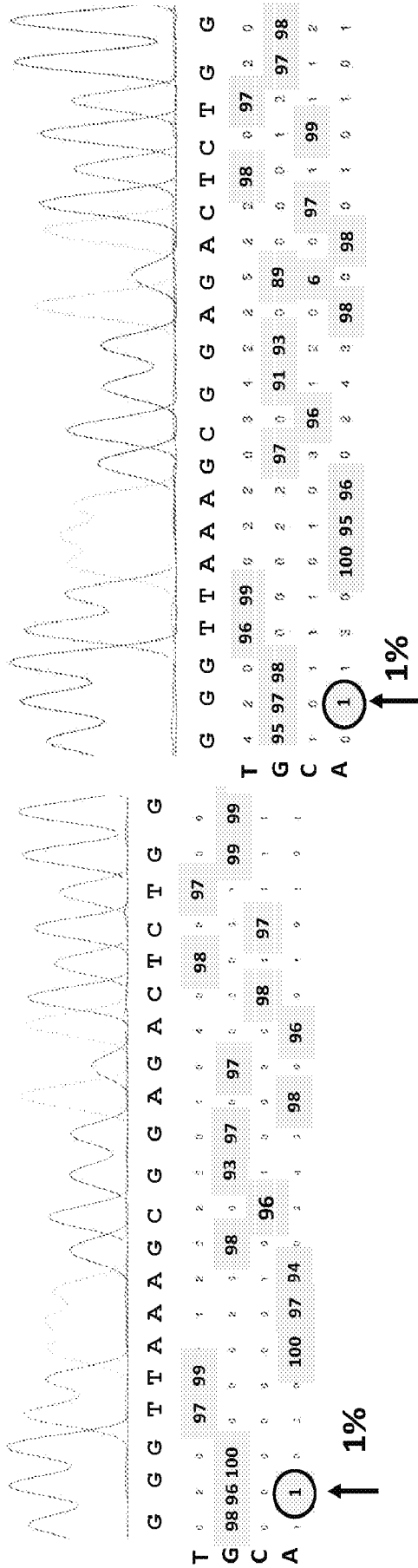
B. Second module target site: 119-U or 151-U:



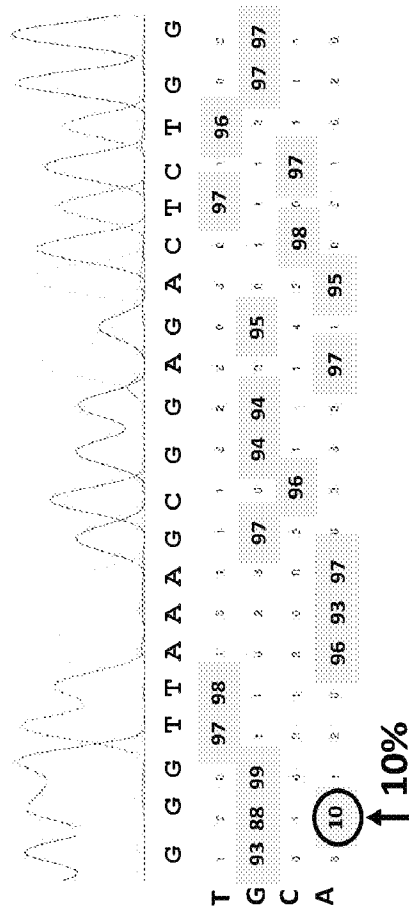
FIGs. 25A, 25B, and 25C



A. Non-treated



C. HEK4 magRNA



FIGS. 26A, 26B, and 26C

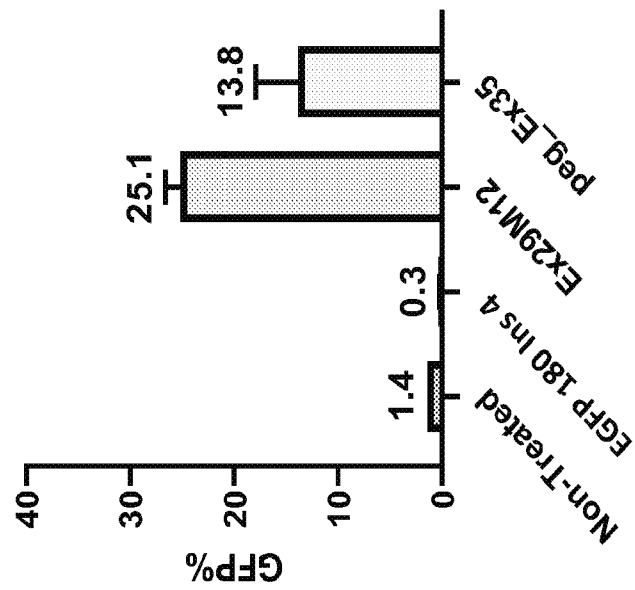
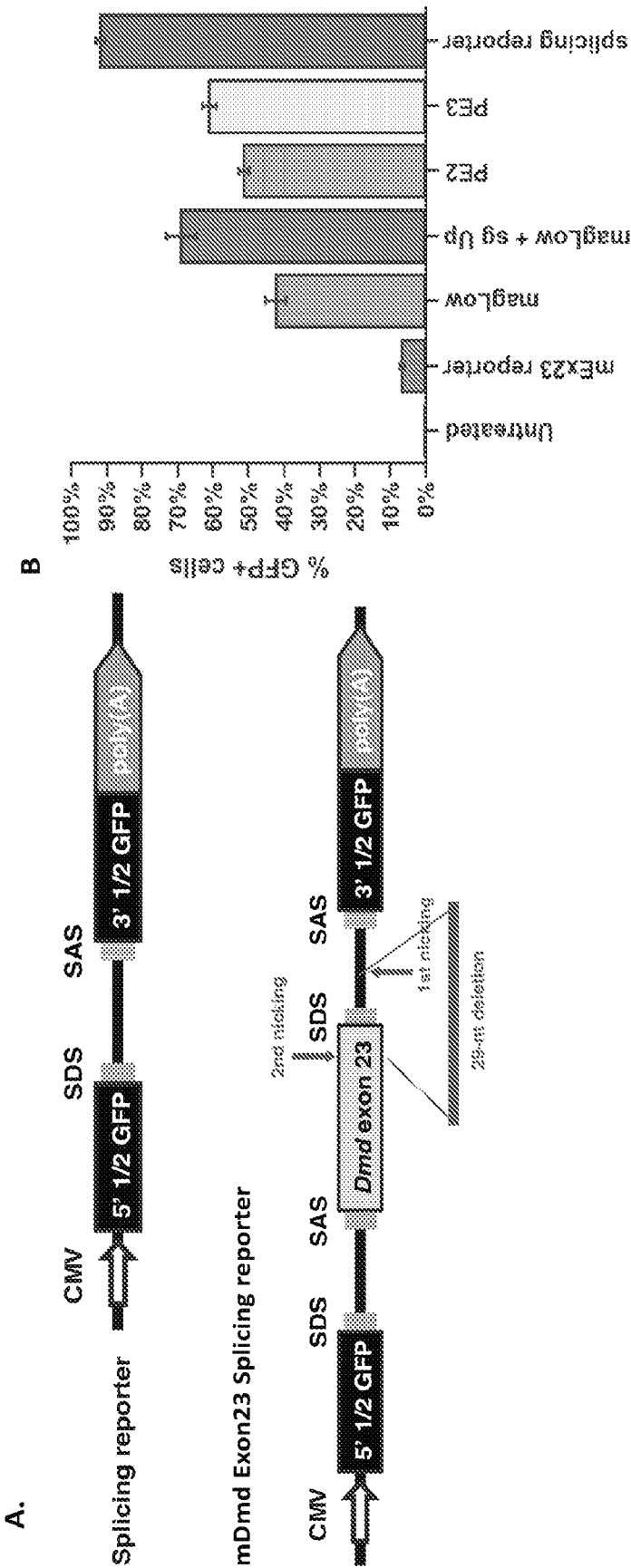


FIG. 27



FIGs. 28A and 28B

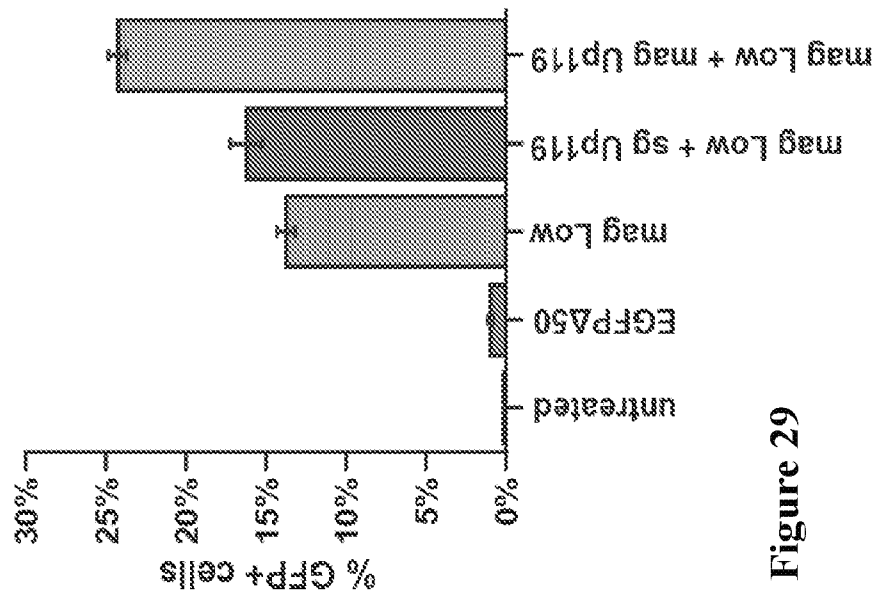
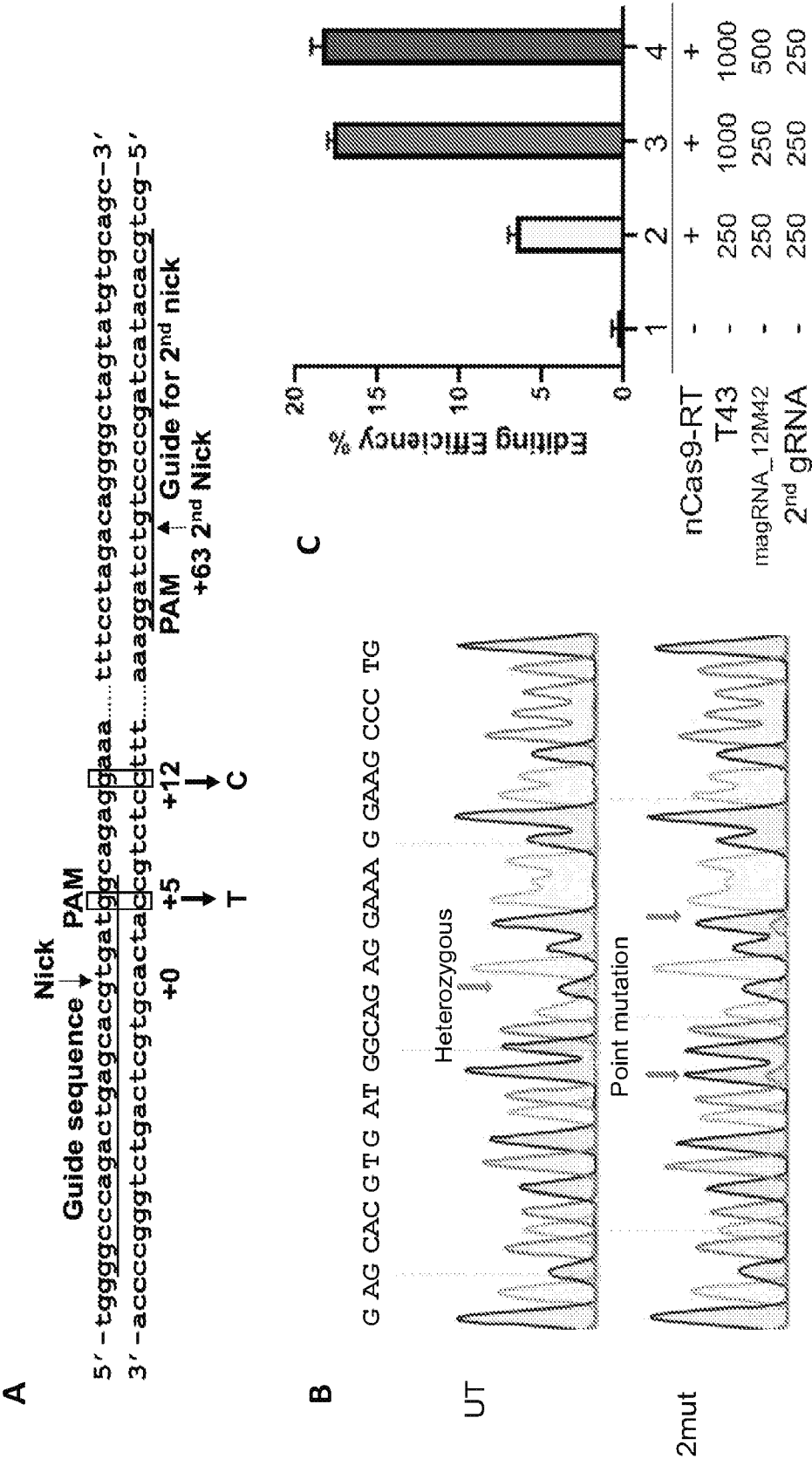
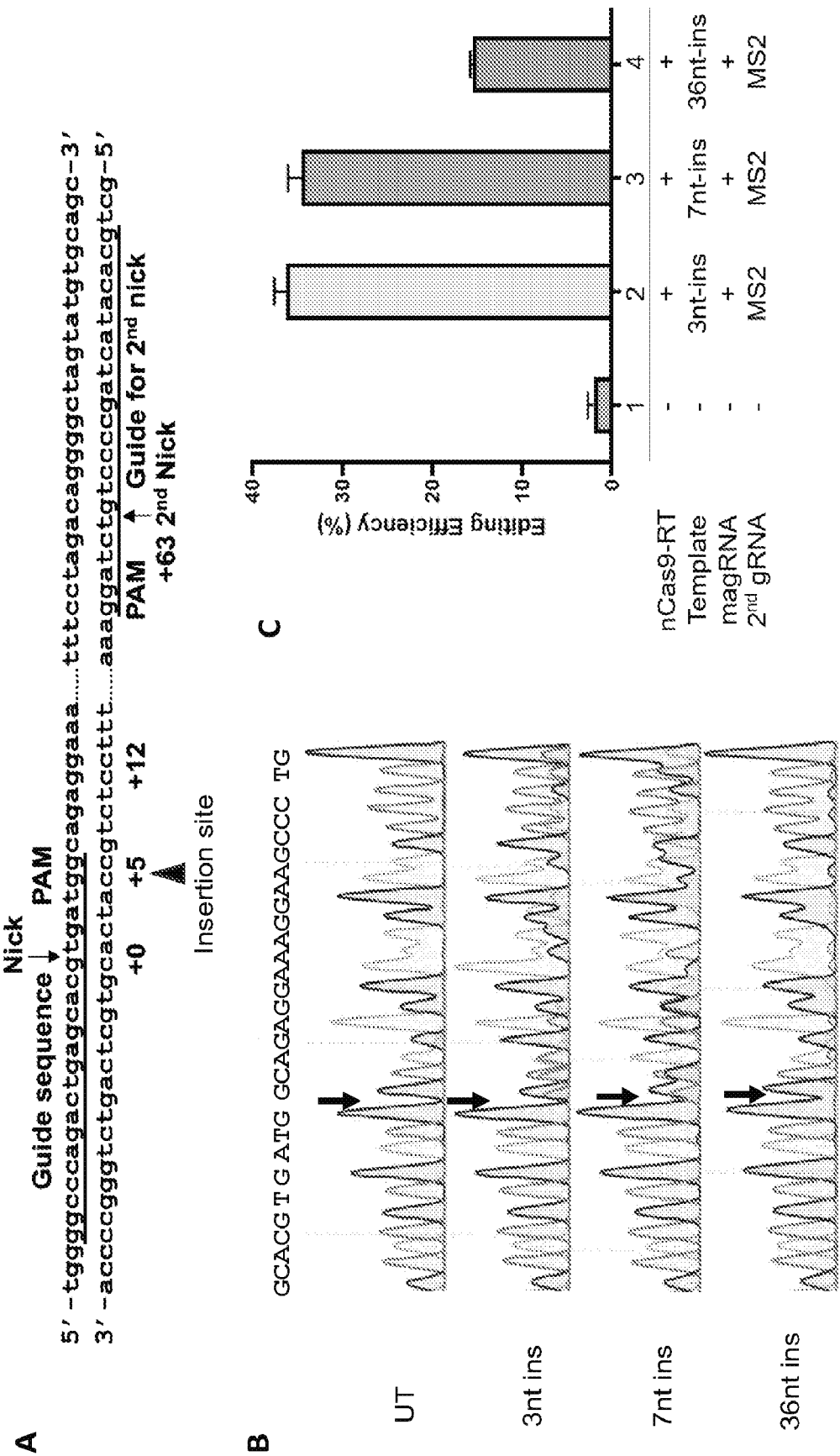


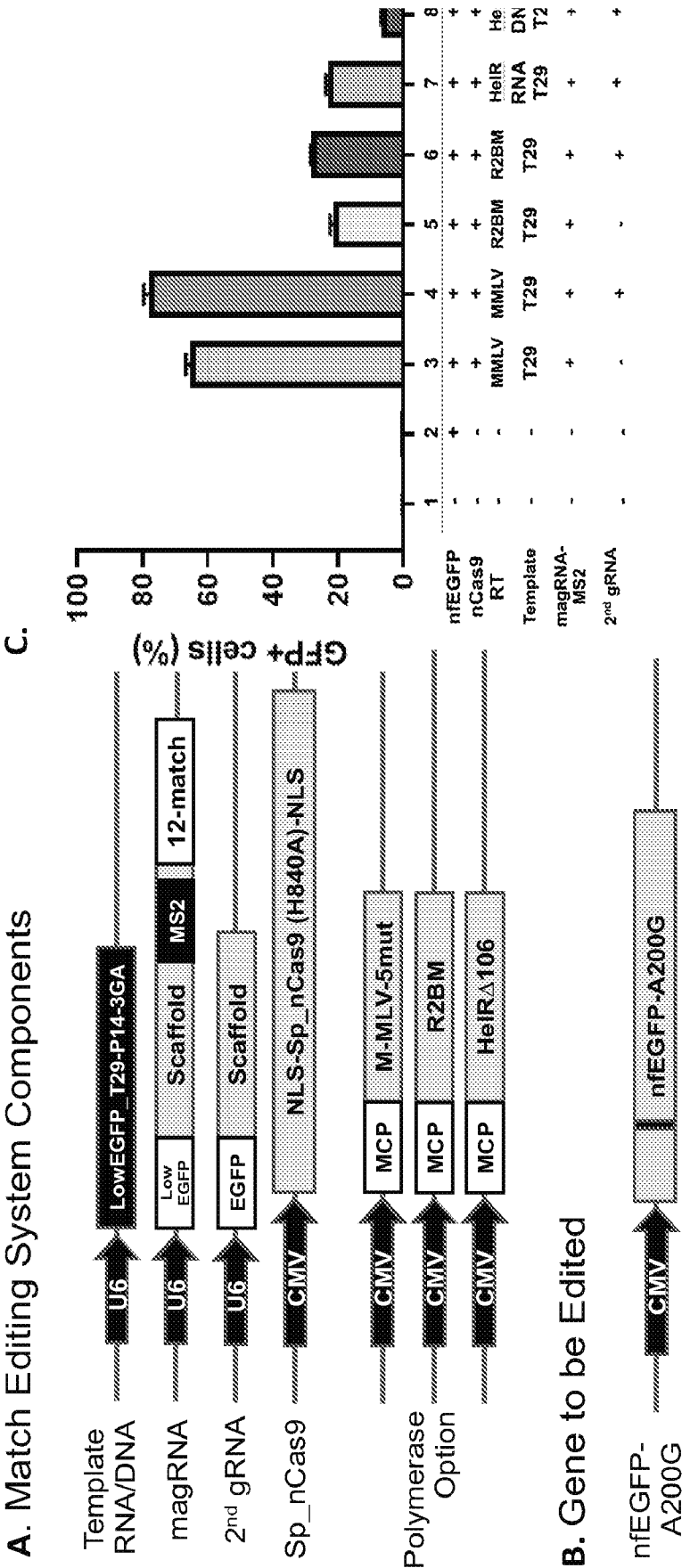
Figure 29



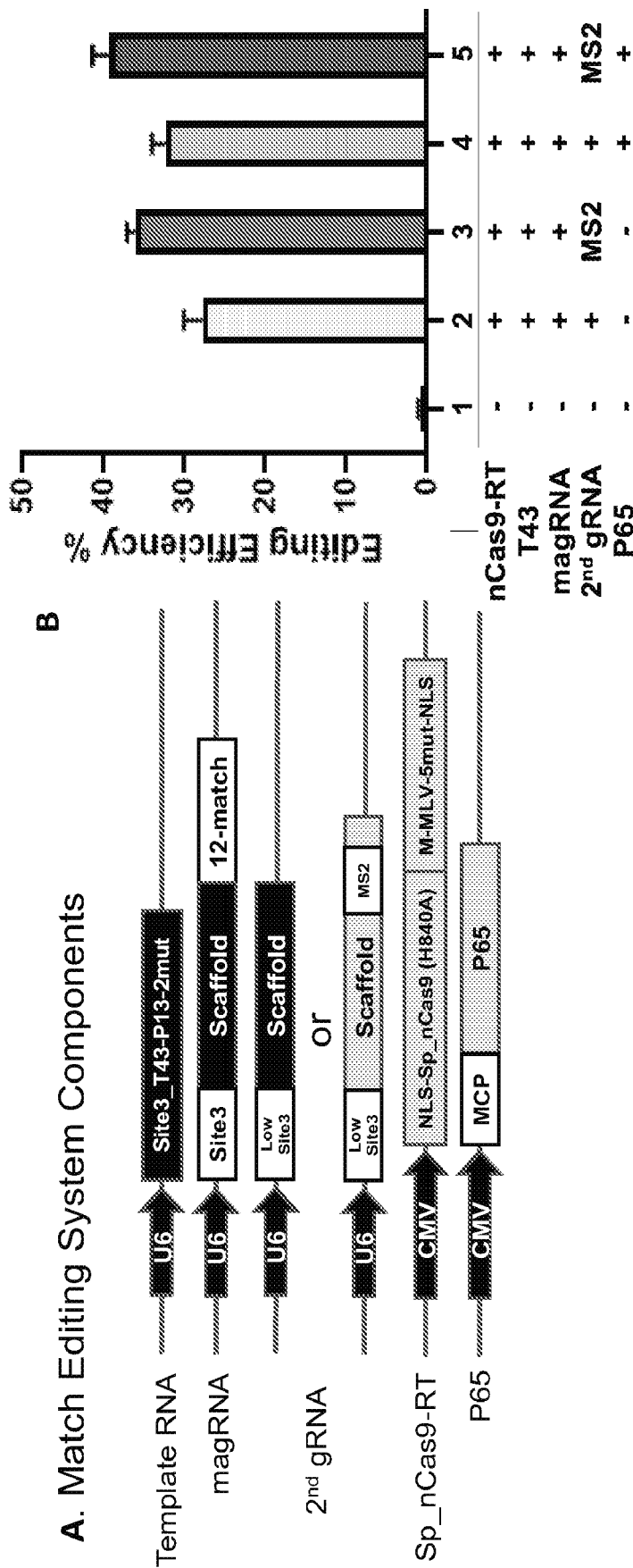
FIGs. 30A, 30B, and 30C



FIGs. 31A, 31B, and 31C

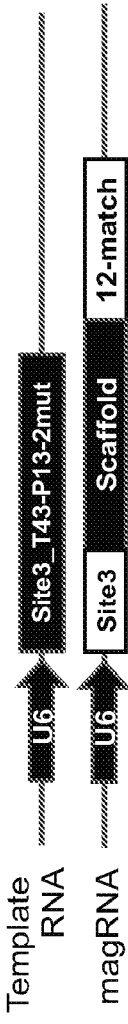


FIGs. 32A, 32B, and 32C



FIGs. 33A and 33B

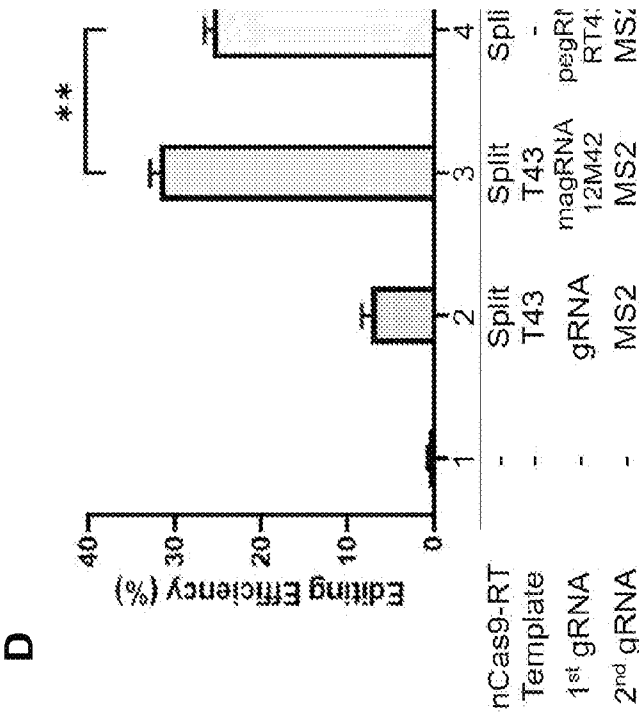
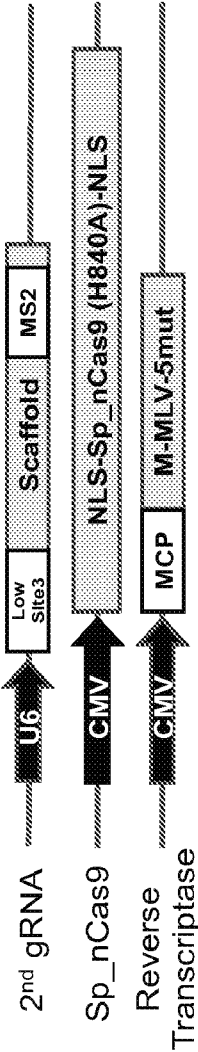
A. ME Specific Components



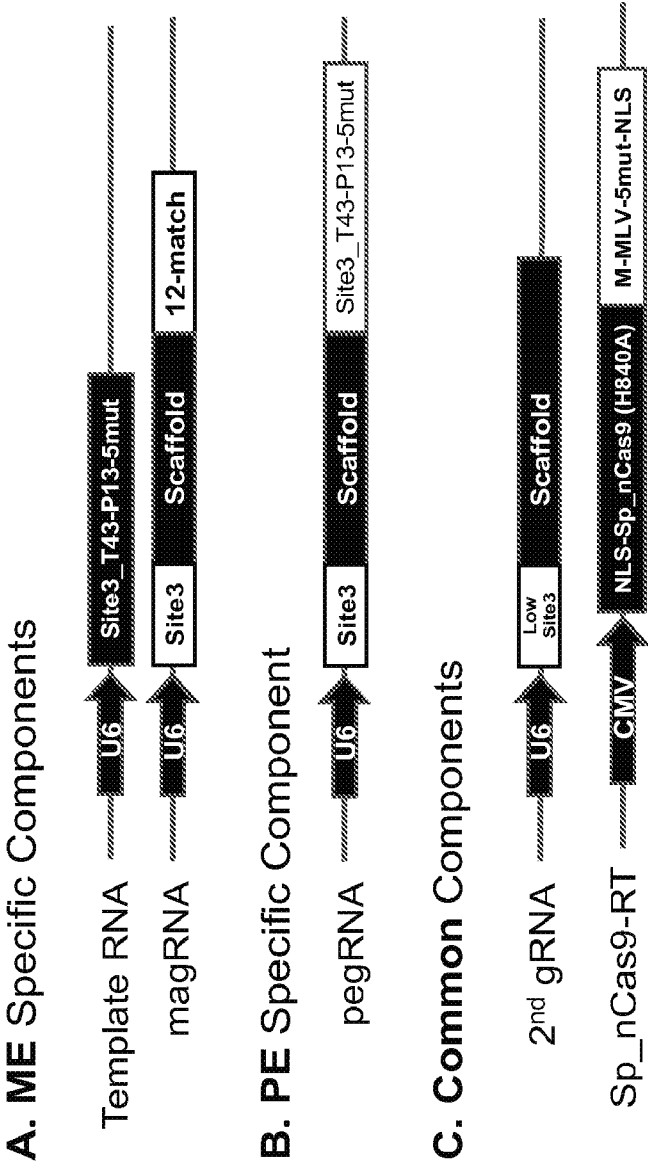
B. PE Specific Component



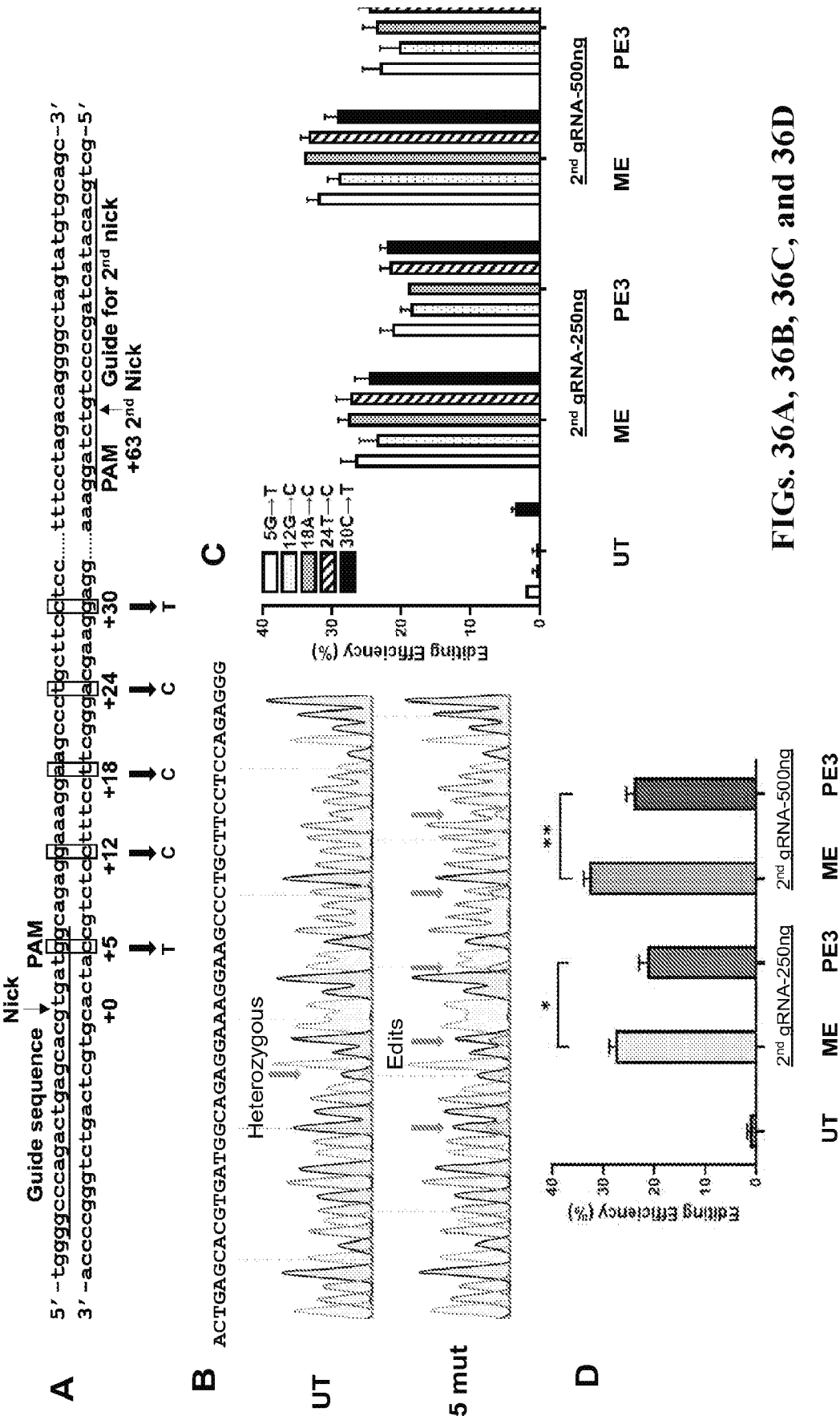
C. Common Components

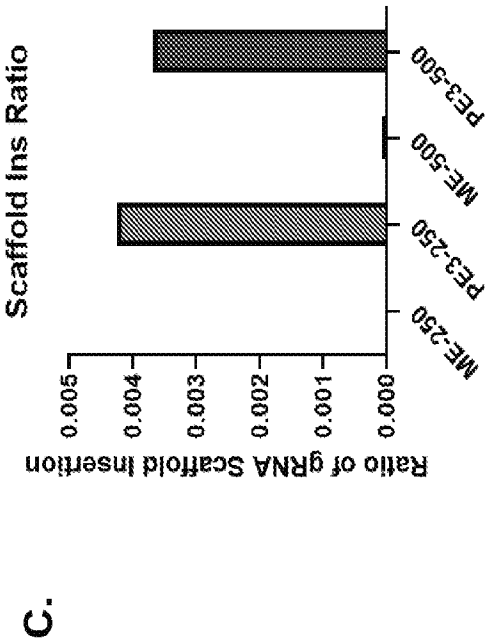
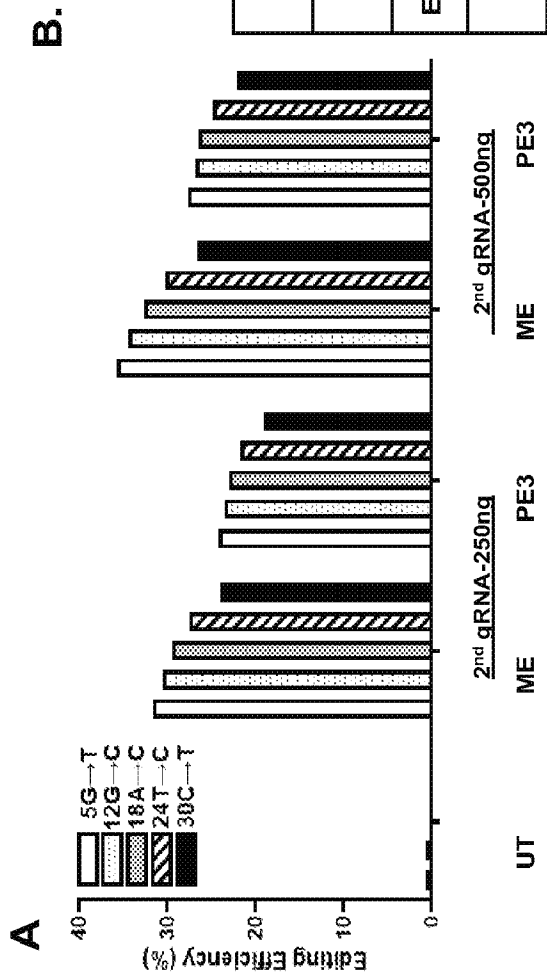


FIGs. 34A, 34B, 34C, and 34D

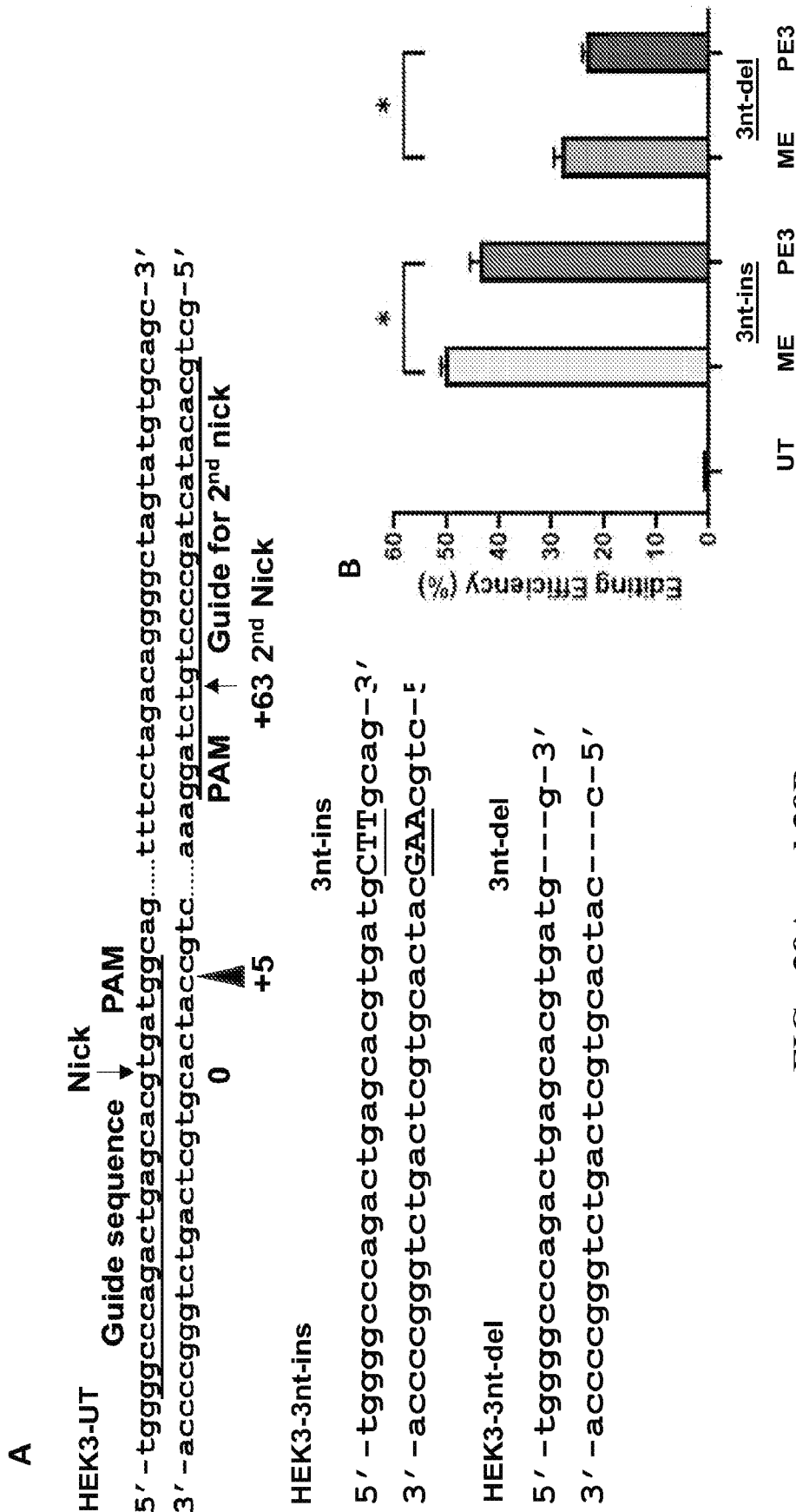


FIGs. 35A, 35B, and 35C

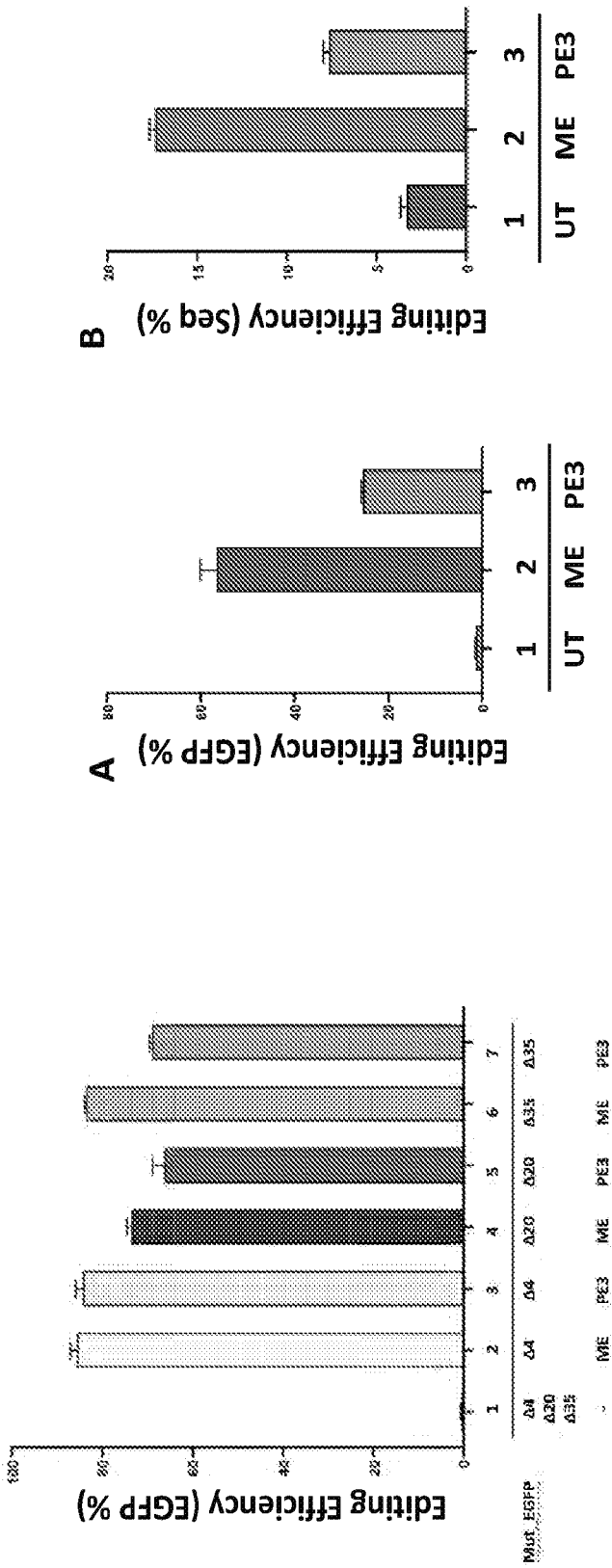




FIGs. 37A, 37B, and 37C

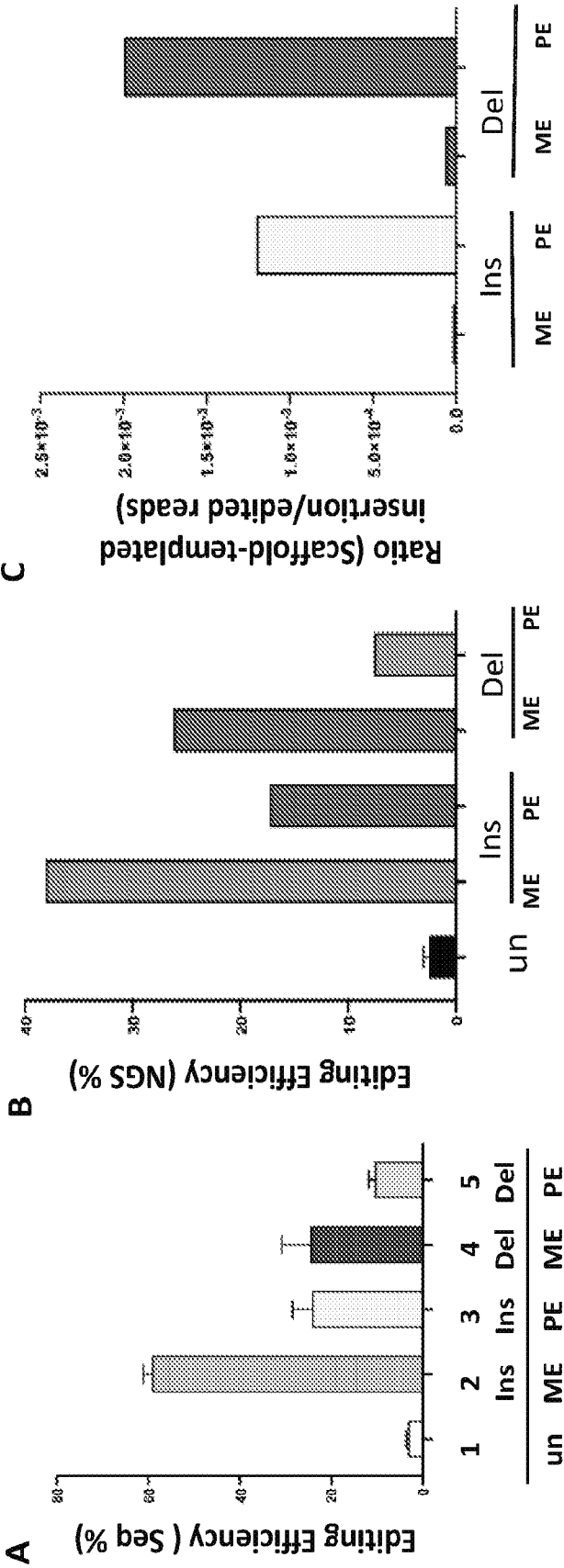


FIGs. 38A and 38B

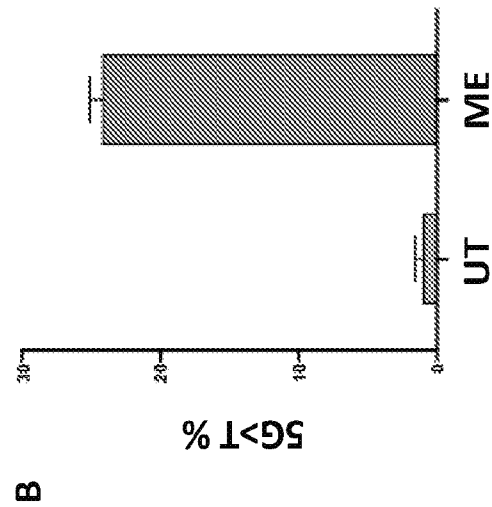
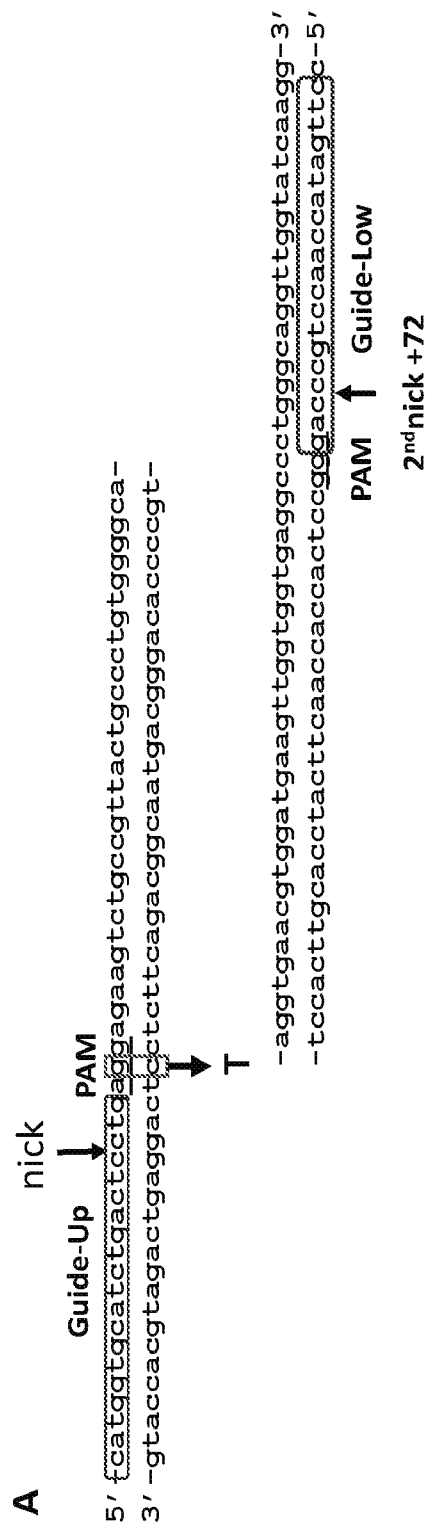


FIGs. 40A and 40B

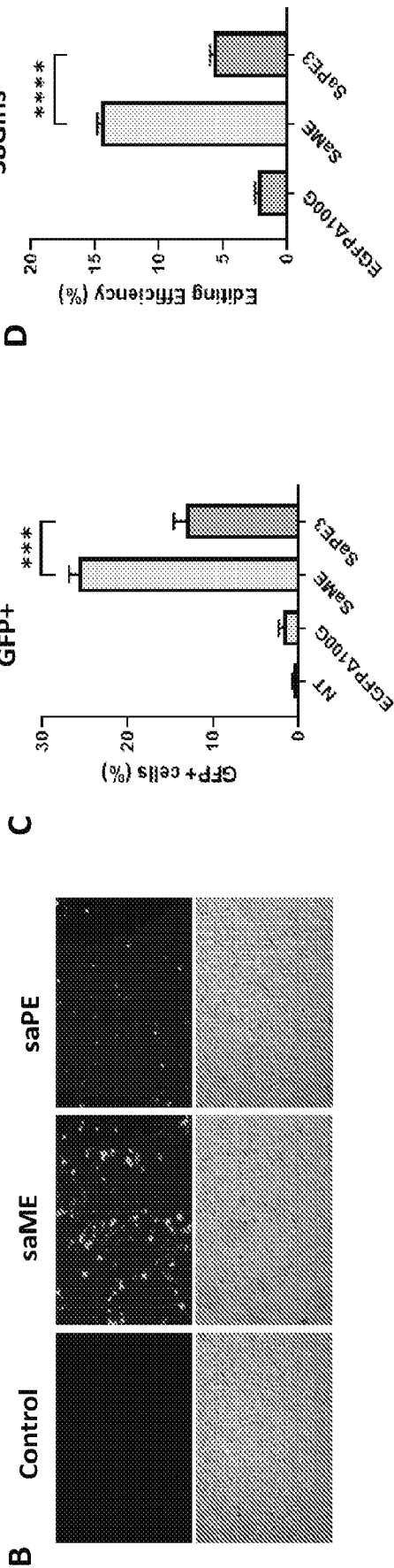
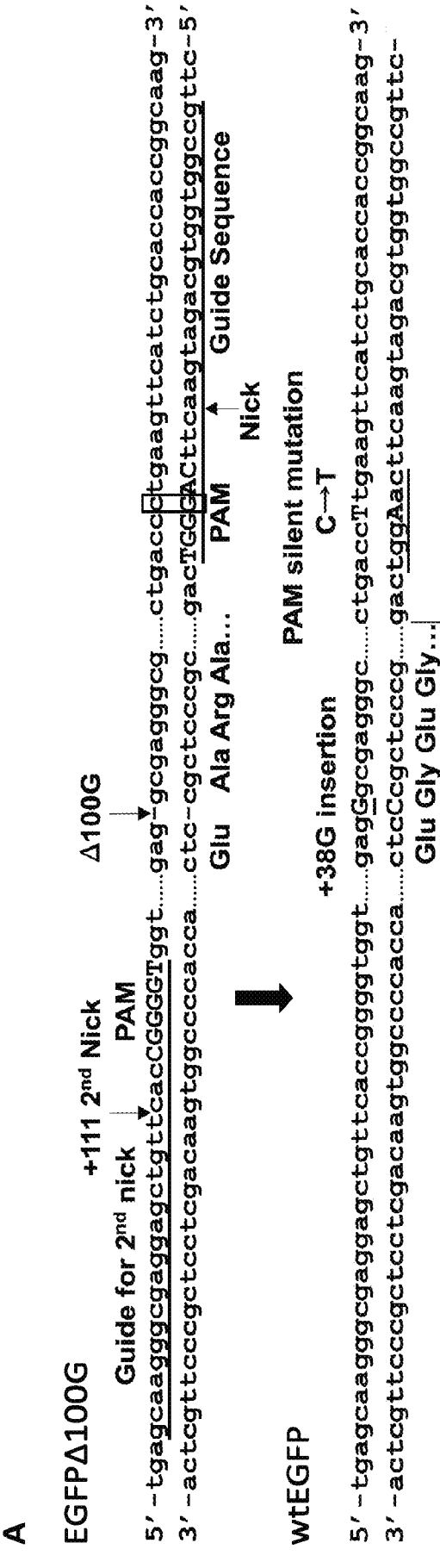
FIG. 39



FIGs. 41A, 41B, and 41C



FIGS. 42A and 42B



FIGs. 43A, 43B, 43C and 43D