

[0038] Figure 1 shows a diagram of prime editing technology as is illustrated in the prior art. nCas9(H840A) = Cas9(H840A nickase; RT = Reverse Transcriptase; PBS = prime binding site.

[0039] Figure 2 shows a diagram of the introduction of a cis-acting regulatory element to the prime editing expression cassette.

[0040] Figure 3 shows PE2-dENE enhances the editing efficiency of PE2. Figure discloses SEQ ID NO: 26.

[0041] Figure 4 shows PE3-dENE enhances the editing efficiency of PE3. Figure discloses SEQ ID NO: 26.

[0042] Figure 5 shows 3'-UTR dENE improves PE editing efficiency on HEK3 target in K562 cells.

[0043] Figure 6 shows 3'-UTR dENE does not improve PE editing efficiency on HEK3 target in HEK293 cells.

Detailed Description of the Invention

[0044] Definitions

[0045] Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by a person skilled in the art to which this invention belongs. The following references provide one of skill with a general definition of many of the terms used in this invention: Singleton, *et al.*, Dictionary of Microbiology and Molecular Biology (2nd ed. 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988); The Glossary of Genetics, 5th Ed., R. Rieger, *et al.* (eds.), Springer Verlag (1991); and Hale & Marham, The Harper Collins Dictionary of Biology (1991). As used herein, the following terms have the meanings ascribed to them unless specified otherwise.

[0046] When introducing elements of the present disclosure or the preferred embodiments(s) thereof, the articles "a," "an," "the" and "said" are intended to mean that there are one or more of the elements. The terms "comprising," "including" and "having" are intended to be inclusive and mean that there may be additional elements other than the listed elements.

[0047] The transitional phrases "comprising," "consisting essentially of" and "consisting of" have the meanings as given in MPEP 2111.03 (Manual of Patent Examining Procedure; United States Patent and Trademark Office). Any claims using the transitional phrase "consisting essentially of" will be understood as reciting only

the transcriptase template of the pegRNA; Equilibration between the edited 3' flap and the unedited 5' flap, endogenous cellular 5' flap cleavage and ligation, and DNA repair results in stably edited DNA (Anzalone *et al.*, 2019; ref. Figure 1). To date, several versions of prime editors (PEs) have been developed. PE1 [SEQ ID NO: 1] is designated by using wild-type Moloney murine leukemia virus reverse transcriptase (M-MLV RT) fused to the C terminus of Cas9(H840A) nickase. PE2 [SEQ ID NO: 2] uses an engineered M-MLV RT. PE3 is defined by introducing an additional guide RNA to nick the non-edited strand, which increases editing efficiency, albeit indel frequency. In PE3b (Anzalone, *et al.*), this nicking single guide RNA (sgRNA) targets the edited sequence, thereby preventing nicking of the non-edited strand until after editing occurs, resulting in fewer indels in mammalian cells.

[0061] Cis-Regulatory Elements

[0062] The present invention substantially improves the efficient of the prime editing system by incorporating one or more *Cis*-regulatory elements (CREs) into the system. (see, Figure 2) One of ordinary skill in the art will understand, armed with the teachings of this specification, that *Cis*-regulatory elements other than the ones specifically exemplified herein may also be suitable use with the present invention and that it is within the skill and knowledge of the ordinary practitioner of the art, following the teaching of the present specification, to screen said *Cis*-regulatory elements without undue experimentation.

[0063] As Wittkapp and Kalay teach us (Nature Reviews Genetics, Jan. 2012, vol. 13, pp. 59 – 69), *Cis*-regulatory element(s) is a term for a collection of transcription factor binding sites and other non-coding DNA that are sufficient to activate (or inhibit) transcription in a defined spatial and/or temporal expression domain. *Cis*-regulatory elements are a type of *Cis*-regulatory sequence needed to activate and sustain transcription. They are composed of DNA (typically, non-coding DNA) containing binding sites for transcription factors and other regulatory molecules. Promoters, enhancers and silencers are the most commonly recognized types of CREs.

[0064] Promoters are required for transcription in eukaryotes but, typically, produce only basal levels of mRNA. Enhancers are more variable than promoters and aid in the upregulation of expression and transcription.

[0085] The RNA-guided endonuclease may comprise at least one nuclear localization signal. In general, an NLS comprises a stretch of basic amino acids. Nuclear localization signals are known in the art (see, *e.g.*, Lange *et al.*, J. Biol. Chem., 2007, 282:5101-5105). For example, in one embodiment, the NLS can be a monopartite sequence, such as PKKKRKV (SEQ ID NO:19) or PKKKRRV (SEQ ID NO:3). In another embodiment, the NLS can be a bipartite sequence. In still another embodiment, the NLS can be KRPAATKKAGQAKKKK (SEQ ID NO:20). The NLS can be located at the N-terminus, the C-terminal, or in an internal location of the RNA-guided endonuclease.

[0086] In some embodiments, the RNA-guided endonuclease can further comprise at least one cell-penetrating domain. In one embodiment, the cell-penetrating domain can be a cell-penetrating peptide sequence derived from the HIV-1 TAT protein. As an example, the TAT cell-penetrating sequence can be GRKKRRQRRPPQPKKKRKV (SEQ ID NO:21). In another embodiment, the cell-penetrating domain can be TLM (PLSSIFSRIGDPPKKRKV; SEQ ID NO:22), a cell-penetrating peptide sequence derived from the human hepatitis B virus. In still another embodiment, the cell-penetrating domain can be MPG (GALFLGWLGAAGSTMGAPKKRKV; SEQ ID NO:23 or GALFLGFLGAAGSTMGAWSQPKKKRKV; SEQ ID NO:24). In an additional embodiment, the cell-penetrating domain can be Pep-1 (KETWWETWWTEWSQPKKKRKV; SEQ ID NO:25), VP22, a cell penetrating peptide from Herpes simplex virus, or a polyarginine peptide sequence. The cell-penetrating domain can be located at the N-terminus, the C-terminus, or in an internal location of the protein.

[0087] In still other embodiments, the RNA-guided endonuclease can also comprise at least one marker domain. Non-limiting examples of marker domains include fluorescent proteins, purification tags, and epitope tags. In some embodiments, the marker domain can be a fluorescent protein. Non limiting examples of suitable fluorescent proteins include green fluorescent proteins (*e.g.*, GFP, GFP-2, tagGFP, turboGFP, EGFP, Emerald, Azami Green, Monomeric Azami Green, CopGFP, AceGFP, ZsGreen1), yellow fluorescent proteins (*e.g.*, YFP, EYFP, Citrine, Venus, YPet, PhiYFP, ZsYellow1), blue fluorescent proteins (*e.g.*, EBFP, EBFP2, Azurite, mKalama1, GFPuv, Sapphire, T-sapphire), cyan fluorescent proteins (*e.g.*, ECFP, Cerulean, CyPet, AmCyan1, Midoriishi-Cyan), red fluorescent proteins (mKate, mKate2, mPlum, DsRed monomer, mCherry, mRFP1, DsRed-Express, DsRed2, DsRed-Monomer,

HcRed-Tandem, HcRed1, AsRed2, eqFP611, mRaspberry, mStrawberry, Jred), and orange fluorescent proteins (mOrange, mKO, Kusabira-Orange, Monomeric Kusabira-Orange, mTangerine, tdTomato) or any other suitable fluorescent protein. In other embodiments, the marker domain can be a purification tag and/or an epitope tag. Exemplary tags include, but are not limited to, glutathione-S-transferase (GST), chitin binding protein (CBP), maltose binding protein, thioredoxin (TRX), poly(NANP), tandem affinity purification (TAP) tag, myc, AcV5, AU1, AU5, E, ECS, E2, FLAG, HA, nus, Softag 1, Softag 3, Strep, SBP, Glu-Glu, HSV, KT3, S, S1, T7, V5, VSV-G, 6 x His (SEQ ID NO: 29), biotin carboxyl carrier protein (BCCP), and calmodulin.

[0088] In certain embodiments, the RNA-guided endonuclease may be part of a protein-RNA complex comprising a pegRNA. The pegRNA interacts with the RNA-guided endonuclease to direct the endonuclease to a specific target site, wherein the 5' end of the guide RNA base pairs with a specific protospacer sequence.

[0089] (II) Fusion Proteins

[0090] Another aspect of the present disclosure provides a fusion protein comprising a CRISPR/Cas-like protein or fragment thereof and an effector domain in combination with a pegRNA and a *cis*-acting regulatory element. The CRISPR/Cas-like protein is directed to a target site by a pegRNA, at which site the effector domain can modify or effect the targeted nucleic acid sequence. The effector domain can be a cleavage domain, an epigenetic modification domain, a transcriptional activation domain, or a transcriptional repressor domain. The fusion protein can further comprise at least one additional domain chosen from a nuclear localization signal, a cell-penetrating domain, or a marker domain.

[0091] (a) CRISPR/Cas-Like Protein

[0092] The fusion protein comprises a CRISPR/Cas-like protein or a fragment thereof. CRISPR/Cas-like proteins are detailed above in section (I). The CRISPR/Cas-like protein can be located at the N-terminus, the C-terminus, or in an internal location of the fusion protein

[0093] In some embodiments, the CRISPR/Cas-like protein of the fusion protein can be derived from a Cas9 protein. The Cas9-derived protein can be wild type, modified, or a fragment thereof. In some embodiments, the Cas9-derived protein can be modified to contain only one functional nuclease domain (either a RuvC-like or a HNH-like nuclease domain). For example, the Cas9-derived protein can be

Claims

- 1) A synthetic nucleic acid composition comprising: i) a sequence encoding a CRISPR-Cas protein, ii) a sequence encoding a reverse transcriptase, and iii) a sequence encoding a cis-acting regulatory element.
- 2) The synthetic nucleic acid composition of Claim 1, wherein said CRISPR-Cas protein is nCas9-H840A.
- 3) The synthetic nucleic acid composition of any of Claims 1 or 2, wherein said reverse transcriptase is M-MLV-RT.
- 4) The synthetic nucleic acid composition of any of Claims 1 – 3, wherein said cis-acting regulatory element is dENE, ENE or sRSM1.
- 5) The synthetic nucleic acid composition of any of Claim 1 - 4, wherein said nucleic acid is DNA or RNA.
- 6) The synthetic nucleic acid composition of any of Claims 1 – 5, wherein said composition further comprises an expression promotor.
- 7) The synthetic nucleic acid composition of Claim 6, wherein said composition is in an expression vector or is incorporated into a transfection virus.
- 8) The synthetic nucleic acid composition of any of Claims 1 – 7, wherein said cis-acting regulatory element is located after the stop codon of the CRISPR-Cas9 sequence and before an mRNA terminator.
- 9) The synthetic nucleic acid composition of any of Claims 1 – 8 further comprising a prime editing guide RNA (pegRNA), wherein said pegRNA is derived from one of PE1, PE2 and PE2.
- 10) An amino acid sequence encoded by the synthetic nucleic acid composition of any of Claims 1 – 8.

- 11) A method of modifying an endogenous DNA sequence, comprising:
- a) providing: i) an operable expression vector comprising a synthetic nucleic acid composition comprising: 1) a sequence encoding a CRISPR-Cas type II system protein, 2) a sequence encoding a reverse transcriptase, and 3) a sequence comprising a cis-acting regulatory element; ii) a prime editing guide RNA (pegRNA) comprising a prime binding site (PBS); and iii) a cell comprising a target endogenous DNA sequence being at least 50% complementary to the PBS;
 - b) transfecting the cell comprising the endogenous DNA sequence of interest with the synthetic nucleic acid composition and pegRNA of the present invention; and
 - c) culturing said transfected cell such that the desired modification is made to the endogenous DNA sequence.
- 12) The method of Claim 11, wherein said CRISPR-Cas type II system protein is a Cas9 protein.
- 13) The method of Claim 11, wherein said endogenous DNA sequence is
- a) at least 75% complementary to the PBS;
 - b) at least 90% complementary to the PBS;
 - c) at least 95% complementary to the PBS;
 - d) at least 98% complementary to the PBS; or
 - e) 100% complementary to the PBS.
- 14) The method of any of Claims 11 – 13, wherein said CRISPR-Cas protein is nCas9-H840A.
- 15) The method of any of Claims 11 – 14, wherein said reverse transcriptase is M-MLV-RT.
- 16) The method of any of Claims 11 – 15, wherein said cis-acting regulatory element is Dene, ENE or sRSM1.

- 17) The method of any of Claims 11 – 16, wherein said operable expression vector is DNA or RNA.
- 18) The method of any of Claims 11 – 17, wherein said composition is incorporated into a transfection virus.
- 19) The method of any of Claims 11 - 18, wherein said cis-acting regulatory element is located after the stop codon of the CRISPR-Cas9 sequence and before an mRNA terminator.
- 20) The method of any of Claims 11 - 19, wherein said pegRNA is derived from one of PE1, PE2 and PE3.
- 21) The method of any of Claims 11 – 20, wherein said CRISPR/Cas type II system protein is introduced into the cell encoded in an operable expression vector.