AAV DELIVERY OF NUCLEOBASE EDITORS

RELATED APPLICATIONS

[0001] This application claims priority under 35 U.S.C. § 119(e) to U.S. provisional applications, U.S.S.N. 62/408,575, filed October 14, 2016, and U.S.S.N. 62/475,780, filed March 23, 2017, each of which is incorporated herein by reference.

GOVERNMENT SUPPORT

[0002] This invention was made with government support under grant number R35 GM118062 and R01 EB022376 awarded by the National Institutes of Health. The government has certain rights in the invention.

BACKGROUND

[0003] Precise genome targeting technologies using the CRISPR/Cas9 system have recently been explored in a wide range of applications, including gene therapy. A major limitation to the application of Cas9 and Cas9-based genome-editing agents in gene therapy is the size of Cas9 (>4 kb), impeding its efficient delivery via recombinant adeno-associated virus (rAAV).

SUMMARY

[0004] Described herein are systems, compositions, kits, and methods for delivering a Cas9 protein or a nucleobase editor to cells, *e.g.*, via recombinant adeno-associated virus vectors. Typically a Cas9 protein or a nucleobase editor is "split" into an N-terminal portion and a C-terminal portion. The N-terminal portion or C-terminal portion of a Cas9 protein or a nucleobase editor may be fused to one member of the intein system, respectively. The resulting fusion proteins, when delivered on separate vectors (*e.g.*, separate rAAV vectors) into one cell and co-expressed, may be joined to form a complete and functional Cas9 protein or nucleobase editor (*e.g.*, via intein-mediated protein splicing). Further provided herein are empirical testing of regulatory elements in the delivery vectors for high expression levels of the split Cas9 protein or the nucleobase editor.

[0005] Some aspects of the present disclosure provide compositions comprising: (i) a first nucleotide sequence encoding a N-terminal portion of a Cas9 protein fused at its C-terminus to an intein-N; and (ii) a second nucleotide sequence encoding an intein-C fused to the N-terminus

of a C-terminal portion of the Cas9 protein, wherein the first nucleotide sequence or second nucleotide sequence is operably linked to a nucleotide sequence encoding at least one bipartite nuclear localization signal.

[0006] In some embodiments, the N-terminal portion of the Cas9 protein comprises a portion of any one of SEQ ID NOs: 1-275 and 394-397 that corresponds to amino acids 1-573 or 1-637 of SEQ ID NO: 1. In some embodiments, the C-terminal portion of the Cas9 protein comprises a portion of any one of SEQ ID NOs: 1-275 and 394-397 that corresponds to amino acids 574-1368 or 638-1368 of SEQ ID NO: 1. In some embodiments, the intein-N comprises the amino acid sequence as set forth in SEQ ID NO: 350-351 and 354-355. In some embodiments, the intein-C comprises the amino acid sequence as set forth in SEQ ID NO: 352-353 and 356-357.

[0007] In some embodiments, the first nucleotide sequence or the second nucleotide sequence further comprises a nucleotide encoding a guide RNA (gRNA) operably linked to a promoter.

[0008] In some embodiments, the first nucleotide sequence or the second nucleotide sequence further comprises a transcriptional terminator. In some embodiments, the transcriptional terminator is the transcriptional terminator from a bGH gene, hGH gene, or SV40 gene. In some embodiments, the transcriptional terminator is the transcriptional terminator from a bGH gene.

[0009] In some embodiments, the first nucleotide sequence or the second nucleotide sequence further comprises a woodchuck hepatitis posttranscriptional regulatory element (WPRE) inserted 5´ of the transcriptional terminator.

[0010] In some embodiments, the bipartite nuclear localization signal comprises an amino acid sequence selected from the group consisting of: KRPAATKKAGQAKKKK (SEQ ID NO: 344), KKTELQTTNAENKTKKL(SEQ ID NO: 345), KRGINDRNFWRGENGRKTR(SEQ ID NO: 346), and RKSGKIAAIVVKRPRK(SEQ ID NO: 347). In some embodiments, the bipartite nuclear localization signal comprises the amino acid sequence as set forth in SEQ ID NO: 344.

[0011] In some embodiments, the Cas9 protein is a catalytically inactive Cas9 (dCas9) or a Cas9 nickase (nCas9), and wherein the first nucleotide sequence of (i) further comprises a nucleotide sequence encoding a nucleobase modifying enzyme fused to the N-terminus of the N-terminal portion of the Cas9 protein.

[0012] In some embodiments, the Cas9 protein is a catalytically inactive Cas9 (dCas9) or a Cas9 nickase (nCas9), and wherein the second nucleotide sequence of (ii) further comprises a

nucleotide sequence encoding a nucleobase modifying enzyme fused to the C-terminus of the C-terminal portion of the Cas9 protein.

[0013] In some embodiments, the nucleobase modifying enzyme is a deaminase. In some embodiments, the deaminase is a cytosine deaminase. In some embodiments, the deaminase is an adenosine deaminase. In some embodiments, the second nucleotide sequence of (ii) further comprises a nucleotide sequence encoding a uracil glycosylase inhibitor (UGI) fused at the 3′ end of the second nucleotide sequence. In some embodiments, the first nucleotide sequence of (i) further comprises a nucleotide sequence encoding a uracil glycosylase inhibitor (UGI) at the 5′ end of the first nucleotide sequence. In some embodiments, the UGI comprises the amino acids sequence of SEQ ID NOs: 299-302.

[0014] In some embodiments, the first nucleotide sequence and the second nucleotide sequence are on different vectors. In some embodiments, the each of the different vectors is a genome of a recombinant adeno-associated virus (rAAV). In some embodiments, each vector is packaged in a rAAV particle.

[0015] Other aspects of the present disclosure provide compositions comprising: (i) a first recombinant adeno associated virus (rAAV) particle comprising a first nucleotide sequence encoding a N-terminal portion of a Cas9 protein fused at its C-terminus to an intein-N; and (ii) a second recombinant adeno associated virus (rAAV) particle comprising a second nucleotide sequence encoding an intein-C fused to the N-terminus of a C-terminal portion of the Cas9 protein, wherein the first nucleotide sequence or second nucleotide sequence is operably linked to a nucleotide sequence encoding at least one bipartite nuclear localization sign.

[0016] Cells comprising the compositions described herein are provided. In some embodiments, the N-terminal portion of the Cas9 protein and the C-terminal portion of the Cas9 protein are joined together to form the Cas9 protein. In some embodiments, the cell is a prokaryotic cell. In some embodiments, the cell is a bacterial cell. In some embodiments, the cell is a eukaryotic cell. In some embodiments, the cell is a yeast cell, a plant cell, or a mammalian cell. In some embodiments, the cell is a human cell.

[0017] Further provided herein are kits comprising the any of the compositions described herein.

[0018] Some aspects of the present disclosure provide compositions comprising: (i) a first nucleotide sequence encoding a N-terminal portion of a nucleobase editor fused at its C-terminus

to an intein-N; and (ii) a second nucleotide sequence encoding an intein-C fused to the N-terminus of a C-terminal portion of the nucleobase editor.

[0019] In some embodiments, the intein-N comprises the amino acid sequence as set forth in SEQ ID NO: 350-351 and 354-355. In some embodiments, the intein-C comprises the amino acid sequence as set forth in SEQ ID NO: 352-353 and 356-357. In some embodiments, the first nucleotide sequence or the second nucleotide sequence further comprises a nucleotide encoding a guide RNA (gRNA) operably linked to a promoter.

[0020] In some embodiments, the first nucleotide sequence or the second nucleotide sequence further comprises a transcriptional terminator. In some embodiments, the transcriptional terminator is a transcriptional terminator from a bGH gene, hGH gene, or SV40 gene. In some embodiments, the transcriptional terminal is from a bGH gene.

[0021] In some embodiments, the first nucleotide sequence or the second nucleotide sequence further comprises a woodchuck hepatitis posttranscriptional regulatory element (WPRE) inserted 5´ of the transcriptional terminator.

[0022] In some embodiments, the first nucleotide sequence or second nucleotide sequence are operably linked to a nucleotide sequence encoding at least one bipartite nuclear localization signal. In some embodiments, the bipartite nuclear localization signal comprises an amino acid sequence selected from the group consisting of: KRPAATKKAGQAKKKK (SEQ ID NO: 344), KKTELQTTNAENKTKKL(SEQ ID NO: 345), KRGINDRNFWRGENGRKTR(SEQ ID NO: 346), and RKSGKIAAIVVKRPRK(SEQ ID NO: 347). In some embodiments, the bipartite nuclear localization signal comprises the amino acid sequence as set forth in SEQ ID NO: 344.

[0023] In some embodiments, the nucleobase editor comprises a cytosine deaminase fused to the N-terminus of a catalytically inactive Cas9 or a Cas9 nickase. In some embodiments, the cytosine deaminase is selected from the group consisting of: APOBEC1, APOBEC3, AID, and pmCDA1. In some embodiments, the nucleobase editor further comprises a uracil glycosylase inhibitor (UGI). In some embodiments, the UGI comprises the amino acids sequence of SEQ ID NOs: 299-302.

[0024] In some embodiments, the first nucleotide sequence and the second nucleotide sequence are on different vectors. In some embodiments, each of the different vectors is a genome of a recombinant adeno-associated virus (rAAV). In some embodiments, the vector is packaged in a rAAV particle.

[0025] Other aspects of the present disclosure provide compositions comprising: (i) a first recombinant adeno associated virus (rAAV) particle comprising a first nucleotide sequence encoding a N-terminal portion of a nucleobase editor fused at its C-terminus to an intein-N; and (ii) a second recombinant adeno associated virus (rAAV) particle comprising a second nuclei acid encoding an intein-C fused to the N-terminus of a C-terminal portion of the nucleobase editor.

[0026] Cells comprising any of the compositions described herein are provided. In some embodiments, the N-terminal portion of the nucleobase editor and the C-terminal portion of the nucleobase editor are joined together to form the nucleobase editor. In some embodiments, the cell is a prokaryotic cell. In some embodiments, the cell is a bacterial cell. In some embodiments, the cell is an eukaryotic cell. In some embodiments, the cell is a yeast cell, a plant cell, or a mammalian cell. In some embodiments, the cell is a human cell.

[0027] Further provided herein are kits comprising any of the compositions described herein.

[0028] Yet other aspects of the present disclosure provide methods comprising: contacting a cell with any of the compositions described herein, wherein the contacting results in the delivery of the first nucleotide sequence and the second nucleotide sequence into the cell, and wherein the N-terminal portion of the nucleobase editor and the C-terminal portion of the nucleobase editor are joined to form a nucleobase editor.

[0029] Yet other aspects of the present disclosure provide methods comprising: administering to a subject in need there of a therapeutically effective amount of any of the compositions described herein. In some embodiments, the subject has a disease or disorder.

[0030] In some embodiments, the disease or disorder is selected from the group consisting of: cystic fibrosis, phenylketonuria, epidermolytic hyperkeratosis (EHK), chronic obstructive pulmonary disease (COPD), Charcot-Marie-Toot disease type 4J. , neuroblastoma (NB), von Willebrand disease (vWD), myotonia congenital, hereditary renal amyloidosis, dilated cardiomyopathy, hereditary lymphedema, familial Alzheimer's disease, prion disease, chronic infantile neurologic cutaneous articular syndrome (CINCA), and desmin-related myopathy (DRM).

[0031] The details of certain embodiments of the invention are set forth in the Detailed Description of Certain Embodiments, as described below. Other features, objects, and

advantages of the invention will be apparent from the Definitions, Examples, Figures, and Claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0032] The accompanying drawings, which constitute a part of this Application, illustrate several embodiments of the invention and together with the description, serve to explain the principles of the invention.

[0033] *Figures 1A-1C* are graphs showing a "split nucleobase editor" for delivery into cells using recombinant adeno associated virus (rAAV) vectors. *Figure 1A* is a schematic representation of how the nucleobase editor is split into two portions. *Figure 1B* shows that AAV-delivered split nucleobase editor can undergo protein splicing upon expression of the two halves in cells to form a complete nucleobase editor that has comparable activity to a nucleobase editor expressed as a whole. *Figure 1C* shows the formation of a complete nucleobase editor from the two halves via protein splicing mediated by DnaE intein.

[0034] *Figure 2* shows that U1118 cells were efficiently transfected by AAV2 containing nucleic acids encoding mCherry. Different viral titers were tested (2.5-10 µl at 4.5 x 10^{11} vg/ml*) and all resulted in efficient transfection of U118 cells. *vg/ml means viral genome-containing particles per microliter.

[0035] *Figures 3A-3B* are graphs showing high throughput sequence (HTS) results of nucleobase editing by rAAV-delivered split nucleobase editor in U118 and HEK cells. Lipid-transfected nucleobase editor was used as a control. A sgRNA targeting R37 in the PRNP gene was used, and the PRNP gene locus was sequenced. *Figure 3A* shows the HTS reads, and *Figure 3B* summarizes the base editing results.

[0036] *Figure 4* is a graph showing the optimization of the transcriptional terminator used in the AAV constructs encoding the split nucleobase editor. Transcriptional terminators of different sizes and origins were tested. bGH transcriptional terminator is relatively short and efficiently terminates transcription comparably to longer terminator sequences. It was therefore chosen to be used in the downstream experiments.

[0037] *Figures 5A-5B* are graphs showing the results of nucleobase editing with long term (up to 15 days) transduction of AAV encoding the split nucleobase editor in mouse astrocytes expressing human ApoE4 cDNA. The target base is in the codon for arginine 112 and arginine

158 in ApoE4, which is converted to a cysteine upon base editing. *Figure 5A* shows that the editing of arginine 158 increases overtime when the mouse astrocytes were transduced at 10^{10} vg, while editing of arginine 112 remained minimal. The nucleotide sequence 3´ of the codon for arginine 158 sequence features a flanking NGG PAM allowing for high activity by SpCas9 (with guide sequence GAAGCGCCTGGCAGTGTACC, SEQ ID NO: 348), while the nucleotide sequence 3´ of the codon for arginine 112 contains a flanking NAG PAM which does not allow for high activity (with guide sequence GACGTGCGCGGCCGCCTGGTG, SEQ ID NO: 349). *Figure 5B* shows cells transduced with rAAV encoding mCherry at 10^{10} vg (control).

[0038] *Figure 6* is a schematic representation of the optimization of the nuclear localization signal in AAV constructs encoding the split nucleobase editor. The nuclear localization signal controls nuclear import, which must occur for reconstituted nucleobase editor to associate with genomic DNA as a prerequisite for editing, and is a potential rate-limiting step in the process. This schematic shows that the NLS (and NLS optimization) is critical for the nucleobase editor to be imported into the nucleus.

[0039] *Figure 7* is a graph showing the results of base editing using different rAAV split nucleobase editor constructs containing different nuclear localization signals (NLS).

[0040] *Figures 8A-8B* are graphs showing the editing of DNMT1 gene in dissociated mouse cortical neurons using an AAV encoded split nucleobase editor.

[0041] *Figure 9A-9B* are graphs showing the editing of DNMT1 gene in mouse Neuro-2a cell line using either an AAV encoded split nucleobase editor, or a lipid transfected DNA encoded nucleobase editor.

DEFINITIONS

[0042] As used herein and in the claims, the singular forms "a," "an," and "the" include the singular and the plural reference unless the context clearly indicates otherwise. Thus, for example, a reference to "an agent" includes a single agent and a plurality of such agents.

[0043] As used herein, the term "Cas9," "Cas9 protein," or "Cas9 nuclease" refers to an RNA-guided nuclease comprising a Cas9 protein (*e.g.*, Cas9 nucleases from a variety of bacterial species), a fragment, a variant (*e.g.*, a catalytically inactive Cas9 or a Cas9 nickase), or a fusion protein (*e.g.*, a Cas9 fused to another protein domain) thereof. A Cas9 nuclease is also referred to sometimes as a casn1 nuclease or a CRISPR (clustered regularly interspaced short palindromic

repeat)-associated nuclease. CRISPR is an adaptive immune system that provides protection against mobile genetic elements (viruses, transposable elements, and conjugative plasmids). CRISPR clusters contain spacers, sequences complementary to antecedent mobile elements, and target invading nucleic acids. CRISPR clusters are transcribed and processed into CRISPR RNA (crRNA). In type II CRISPR systems correct processing of pre-crRNA requires a trans-encoded small RNA (tracrRNA), endogenous ribonuclease 3 (rnc) and a Cas9 protein. The tracrRNA serves as a guide for ribonuclease 3-aided processing of pre-crRNA. Subsequently, Cas9/crRNA/tracrRNA endonucleolytically cleaves linear or circular dsDNA target complementary to the spacer. In nature, DNA-binding and cleavage typically requires protein and both RNAs. However, single guide RNAs ("sgRNA", or simply "gNRA") can be engineered so as to incorporate aspects of both the crRNA and tracrRNA into a single RNA species. Non-limiting examples of Cas9 proteins and their respective amino acid sequence are provided in Example 1.

[0044] A nuclease-inactive Cas9 protein may interchangeably be referred to as a "dCas9" protein (for nuclease-"dead" Cas9). Methods for generating a Cas9 protein (or a fragment thereof) having an inactive DNA cleavage domain are known (See, *e.g.*, Jinek *et al.*, *Science*. 337:816-821(2012); Qi *et al.*, (2013) *Cell*. 28;152(5):1173-83, incorporated herein by reference). For example, the DNA cleavage domain of Cas9 is known to include two subdomains, the HNH nuclease subdomain and the RuvC1 subdomain. The HNH subdomain cleaves the strand complementary to the gRNA, whereas the RuvC1 subdomain cleaves the non-complementary strand. Mutations within these subdomains can silence the nuclease activity of Cas9. For example, the mutations D10A and H840A completely inactivate the nuclease activity of *S. pyogenes* Cas9 (Jinek *et al.*, *Science*. 337:816-821(2012); Qi *et al.*, *Cell*. 28;152(5):1173-83 (2013). Additional suitable nuclease-inactive dCas9 domains will be apparent to those of skill in the art based on this disclosure and knowledge in the field, and are within the scope of this disclosure. Such additional exemplary suitable nuclease-inactive Cas9 domains include, but are not limited to, D10A/H840A, D10A/D839A/H840A, and D10A/D839A/H840A/N863A mutant domains (See, *e.g.*, Prashant *et al.*, *Nature Biotechnology*. 2013; 31(9):833-838, incorporated herein by reference).

[0045] In some embodiments, a Cas9 nickase is used as part of the nucleobase editor. A Cas9 nickase is able to cleave one strand of the double strand DNA. A Cas9 nickase may be generated

by introducing an inactivating mutation into either the HNH domain or the RuvC1 domain. For example, an inactivating mutation (D10A) may be introduced in the RuvC1 domain of the *S. pyogenes* Cas9, while the HNH domain remains active, *i.e.*, the residue at position 840 remains a histidine. Such Cas9 variants are able to generate a single-strand DNA break (nick) at a specific location based on the gRNA-defined target sequence. One skilled in the art is able to identify the catalytic residues in the RuvC1 and HNH domains of any known Cas9 proteins and introduce inactivating mutations to generate a corresponding dCas9 or nCas9.

[0046] A "split Cas9 protein" or "split Cas9" refers to a Cas9 protein that is provided as an N-terminal portion (also referred to as an N-terminal half) and a C-terminal portion (also referred to as a C-terminal half) encoded by two separate nucleotide sequences. The polypeptides corresponding to the N-terminal portion and the C-terminal portion of the Cas9 protein may be combined (joined) to form a complete Cas9 protein. A Cas9 protein is known to consist of a bi-lobed structure linked by a disordered linker (*e.g.*, as described in Nishimasu *et al.*, *Cell*, Volume 156, Issue 5, pp. 935–949, 2014, incorporated herein by reference). In some embodiments, the "split" occurs between the two lobes, generating two portions of a Cas9 protein, each containing one lobe.

[0047] An "intein" is a segment of a protein that is able to excise itself and join the remaining portions (the exteins) with a peptide bond in a process known as protein splicing. Inteins are also referred to as "protein introns." The process of an intein excising itself and joining the remaining portions of the protein is herein termed "protein splicing" or "intein-mediated protein splicing." In some embodiments, an intein of a precursor protein (an intein containing protein prior to intein-mediated protein splicing) comes from two genes. Such intein is referred to herein as a split intein. For example, in cyanobacteria, DnaE, the catalytic subunit α of DNA polymerase III, is encoded by two separate genes, dnaE-n and dnaE-c. The intein encoded by the dnaE-n gene is herein referred as "intein-N." The intein encoded by the dnaE-c gene is herein referred as "intein-C."

[0048] Other intein systems may also be used. For example, a synthetic intein based on the dnaE intein, the Cfa-N and Cfa-C intein pair, has been described (*e.g.*, in Stevens *et al.*, J Am Chem Soc. 2016 Feb 24;138(7):2162-5, incorporated herein by reference). Non-limiting examples of intein pairs that may be used in accordance with the present disclosure include: Cfa DnaE intein,

Ssp GyrB intein, Ssp DnaX intein, Ter DnaE3 intein, Ter ThyX intein, Rma DnaB intein and Cne Prp8 intein (*e.g.*, as described in US Patent 8,394,604, incorporated herein by reference.

[0049] Exemplary nucleotide and amino acid sequences of inteins are provided.

DnaE Intein-N DNA:
TGCCTGTCATACGAAACCGAGATACTGACAGTAGAATATGGCCTTCTGCCAATCGGG
AAGATTGTGGAGAAACGGATAGAATGCACAGTTTACTCTGTCGATAACAATGGTAA
CATTTATACTCAGCCAGTTGCCCAGTGGCACGACCGGGGAGAGCAGGAAGTATTCG
AATACTGTCTGGAGGATGGAAGTCTCATTAGGGCCACTAAGGACCACAAATTTATG
ACAGTCGATGGCCAGATGCTGCCTATAGACGAAATCTTTGAGCGAGAGTTGGACCTC
ATGCGAGTTGACAACCTTCCTAAT (SEQ ID NO: 350)

DnaE Intein-N Protein:
CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCL
EDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN (SEQ ID NO: 351)

DnaE Intein-C DNA:
ATGATCAAGATAGCTACAAGGAAGTATCTTGGCAAACAAAACGTTTATGATATTGG
AGTCGAAAGAGATCACAACTTTGCTCTGAAGAACGGATTCATAGCTTCTAAT (SEQ
ID NO: 352)
Intein-C:
MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN (SEQ ID NO: 353)

Cfa-N DNA:
TGCCTGTCTTATGATACCGAGATACTTACCGTTGAATATGGCTTCTTGCCTATTGGAA
AGATTGTCGAAGAGAGAATTGAATGCACAGTATATACTGTAGACAAGAATGGTTTC
GTTTACACACAGCCCATTGCTCAATGGCACAATCGCGGCGAACAAGAAGTATTTGA
GTACTGTCTCGAGGATGGAAGCATCATACGAGCAACTAAAGATCATAAATTCATGA
CCACTGACGGGCAGATGTTGCCAATAGATGAGATATTCGAGCGGGGCTTGGATCTC
AAACAAGTGGATGGATTGCCA (SEQ ID NO: 354)

Cfa-N Protein:
CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCL
EDGSIIRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP (SEQ ID NO: 355)

Cfa-C DNA:
ATGAAGAGGACTGCCGATGGATCAGAGTTTGAATCTCCCAAGAAGAAGAGGAAAGT
AAAGATAATATCTCGAAAAAGTCTTGGTACCCAAAATGTCTATGATATTGGAGTGGA
GAAAGATCACAACTTCCTTCTCAAGAACGGTCTCGTAGCCAGCAAC (SEQ ID NO:
356)

Cfa-C Protein:
MKRTADGSEFESPKKKRKVKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN (SEQ ID
NO: 357)

[0050] Intein-N and intein-C may be fused to the N-terminal portion of the split Cas9 and the C-terminal portion of the split Cas9, respectively, for the joining of the N-terminal portion of the split Cas9 and the C-terminal portion of the split Cas9. For example, in some embodiments, an intein-N is fused to the C-terminus of the N-terminal portion of the split Cas9, *i.e.*, to form a structure of N-[N-terminal portion of the split Cas9]-[intein-N]-C. In some embodiments, an intein-C is fused to the N-terminus of the C-terminal portion of the split Cas9, *i.e.*, to form a structure of N-[intein-C]-[C-terminal portion of the split Cas9]-C. The mechanism of intein-mediated protein splicing for joining the proteins the inteins are fused to (*e.g.*, split Cas9) is known in the art, *e.g.*, as described in Shah *et al.*, *Chem Sci.* 2014; 5(1):446–461, incorporated herein by reference.

[0051] Herein, a "nucleobase editor" refers to a protein that edits a nucleotide base. "Edit" refers to the conversion of one nucleobase to another (*e.g.*, A to G, A to C, A to T, C to T, C to G, C to A, G to A, G to C, G to T, T to A, T to C, T to G). In some embodiments, a nucleobase editor is a macromolecule or macromolecular complex that results primarily (*e.g.*, more than 80%, more than 85%, more than 90%, more than 95%, more than 99%, more than 99.9%, or 100%) in the conversion of a nucleobase in a polynucleic acid sequence into another nucleobase (i.e., a transition or transversion) using a combination of 1) a nucleotide-, nucleoside-, or nucleobase-modifying enzyme and 2) a nucleic acid binding protein that can be programmed to bind to a specific nucleic acid sequence.

[0052] In some embodiments, the nucleobase editor comprises a DNA binding domain (*e.g.*, a programmable DNA binding domain such as a dCas9 or nCas9) that directs it to a target sequence. In some embodiments, the nucleobase editor comprises a nucleobase modifying enzyme fused to a programmable DNA binding domain (*e.g.*, a dCas9 or nCas9). A "nucleobase modifying enzyme" is an enzyme that can modify a nucleobase and convert one nucleobase to another (*e.g.*, a deaminase such as a cytosine deaminase or a adenosine deaminase). In some embodiments, the nucleobase editor may target cytosine (C) bases in a nucleic acid sequence and convert the C to thymine (T) base. In some embodiments, the C to T editing is carried out by a deaminase, *e.g.*, a cytosine deaminase. Base editors that can carry out other types of base conversions (*e.g.*, adenosine (A) to guanine (G), C to G) are also contemplated.

[0053] Nucleobase editors that convert a C to T, in some embodiments, comprise a cytosine deaminase. A "cytosine deaminase" refers to an enzyme that catalyzes the chemical reaction

"cytosine + H_2O → uracil + NH_3" or "5-methyl-cytosine + H_2O → thymine + NH_3." As it may be apparent from the reaction formula, such chemical reactions result in a C to U/T nucleobase change. In the context of a gene, such a nucleotide change, or mutation, may in turn lead to an amino acid change in the protein, which may affect the protein's function, *e.g.*, loss-of-function or gain-of-function. In some embodiments, the C to T nucleobase editor comprises a dCas9 or nCas9 fused to a cytosine deaminase. In some embodiments, the cytosine deaminase domain is fused to the N-terminus of the dCas9 or nCas9. In some embodiments, the nucleobase editor further comprises a domain that inhibits uracil glycosylase, and/or a nuclear localization signal. Such nucleobase editors have been described in the art, *e.g.*, in US Patent 9,068,179, US Patent Application Publications US 2015/0166980, published June 18, 2015; US 2015/0166981, published June 18, 2015; US 2015/0166982, published June 18, 2015; US 2015/0166984, published June 18, 2015; and US2015/0165054, published June 18, 2015; and US Provisional Applications, U.S.S.N. 62/245,828, filed October 23, 2015; U.S.S.N. 62/279,346, filed January 15, 2016; U.S.S.N. 62/311,763, filed March 22, 2016; U.S.S.N. 62/322,178, filed April 13, 2016; U.S.S.N. 62/357,352, filed June 30, 2016; U.S.S.N. 62,370,700, filed August 3, 2016; U.S.S.N. 62/398,490, filed September 22, 2016; and U.S.S.N. 62/408,686, filed October 14, 2016; PCT Application PCT/US2016/058344, filed October 22, 2016, US Patent Application, U.S.S.N. 15/311,852, filed October 22, 2016; and in Komor *et al.*, *Nature*, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage, 533, 420–424 (2016), the entire contents of each of which is incorporated herein by reference.

[0054] In some embodiments, a nucleobase editor converts an A to G. In some embodiments, the nucleobase editor comprises an adenosine deaminase. An "adenosine deaminase" is an enzyme involved in purine metabolism. It is needed for the breakdown of adenosine from food and for the turnover of nucleic acids in tissues. Its primary function in humans is the development and maintenance of the immune system. An adenosine deaminase catalyzes hydrolytic deamination of adenosine (forming inosine, which base pairs as G) in the context of DNA. There are no known adenosine deaminases that act on DNA. Instead, known adenosine deaminase enzymes only act on RNA (tRNA or mRNA). Evolved deoxyadenosine deaminase enzymes that accept DNA substrates and deaminate dA to deoxyinosine and here use in adenosine nucleobase editos have been described, *e.g.*, in US provisional application, U.S.S.N. 62/370,684, filed August 3, 2016; US provisional application, U.S.S.N. 62/370,684, filed February 3, 2017, US provisional

application, U.S.S.N. 62/473,714, filed March 20, 2017, and PCT Application PCT/US2017/045381, filed August 3, 2017; each of which is incorporated herein by reference. Non-limiting examples of evolved adenosine deaminases that accept DNA as substrates are provided in Example 1.

[0055] In some embodiments, the adenosine deaminase is *E. coli* TadA (SEQ ID NO: 314). The possible mutations in ecTadA and constructs expressing nucleobase editors comprising the modified ecTadA are provided in Table 2. The sequences of exemplary EcTadA mutants and nucleotibase editors comprising such mutants are provided in Example 1.

Table 2. EcTadA mutants for A to G nucleobase editor

Name	Construct Architecture	Mutations in TadA
pNMG-142	pCMV_ecTadA_XTEN_Cas9n_SGGS_NLS	wild-type
pNMG-143	pCMV_ecTadA_XTEN_Cas9n_SGGS_NLS	D108N
pNMG-144	pCMV_ecTadA_XTEN_Cas9n_SGGS_NLS	A106V_D108N
pNMG-145	pCMV_ecTadA_XTEN_Cas9n_SGGS_NLS	D108G
pNMG-146	pCMV_ecTadA_XTEN_Cas9n_SGGS_NLS	R107C_D108N
pNMG-147	pCMV_ecTadA_XTEN_Cas9n_SGGS_NLS	D108V
pNMG-155	pCMV_ecTadA_XTEN_dead Cas9_SGGS_UGI_NLS	D108N
pNMG-156	pCMV_ecTadA_XTEN_nCas9_SGGS_UGI_SGGS_NLS	D108N
pNMG-157	pCMV_ecTadA_XTEN_dead Cas9_SGGS_UGI_SGGS_NLS	D108G
pNMG-158	pCMV_ecTadA_XTEN_nCas9_SGGS_UGI_SGGS_NLS	D108G
pNMG-160	pCMV_ecTadA_XTEN_nCas9_SGGS_AAG*(E125Q)_SGGS_NLS	D108N
pNMG-161	pCMV_ecTadA_XTEN_Cas9n_SGGS_EndoV*(D35A)_NLS	D108N
pNMG-162	pCMV_ecTadA_XTEN_Cas9n_SGGS_NLS	H8Y_D108N_S127S_D147Y_Q154H
pNMG-163	pCMV_ecTadA_XTEN_Cas9n_SGGS_NLS	H8Y_R24W_D108N_N127S_D147Y_E155V
pNMG-164	pCMV_ecTadA_XTEN_Cas9n_SGGS_NLS	D108N_D147Y_E155V
pNMG-165	pCMV_ecTadA_XTEN_Cas9n_SGGS_NLS	H8Y_D108N_S127S
pNMG-171	pCMV_Cas9n_XTEN_ecTadA_SGGS_NLS	wild-type
pNMG-172	pCMV_Cas9n_XTEN_ecTadA_SGGS_NLS	D108N
pNMG-173	pCMV_Cas9n_XTEN_ecTadA_SGGS_NLS	H8Y_D108N_N127S_D147Y_Q154H

pNMG-174	pCMV_Cas9n_XTEN_ecTadA_SGGS_NLS	H8Y_R24W_D108N_N127S_D147Y_E155V
pNMG-175	pCMV_Cas9n_XTEN_ecTadA_SGGS_NLS	D108N_D147Y_E155V
pNMG-176	pCMV_Cas9n_XTEN_ecTadA_SGGS_NLS	H8Y_D108N_S127S
pNMG-177	pCMV_ecTadA_XTEN_Cas9n_SGGS_NLS	A106V_D108N_D147Y_E155V
pNMG-178	pCMV_ecTadA_XTEN_Cas9n_SGGS_UGI_SGGS_NLS	D108N_D147Y_E155V
pNMG-179	pCMV_ecTadA_XTEN_Cas9n_SGGS_AAG*(E125Q)_SGGS_NLS	A106V_D108N_D147Y_E155V
pNMG-180	pCMV_ecTadA_XTEN_Cas9n_SGGS_UGI_SGGS_NLS	A106V_D108N_D147Y_E155V
pNMG-181	pCMV_ecTadA_XTEN_Cas9n_SGGS_AAG*(E125Q)_SGGS_NLS	D108N_D147Y_E155V
pNMG-182	pCMV_ecTadA_SGGS_nCas9_SGGS_NLS	D108N_D147Y_E155V
pNMG-183	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	D108N_D147Y_E155V
pNMG-235	pCMV_ecTadA_XTEN_Cas9n_XTEN_AAG*(E125A)_SGGS_NLS	A106V_D108N_D147Y_E155V
pNMG-236	pCMV_ecTadA_XTEN_Cas9n_XTEN_AAG*(E125Q)_SGGS_NLS	A106V_D108N_D147Y_E155V
pNMG-237	pCMV_ecTadA_XTEN_Cas9n_XTEN_AAG*(wt)_SGGS_NLS	A106V_D108N_D147Y_E155V
pNMG-238	pCMV_AAG*(E125A)_XTEN_ecTadA_XTEN_Cas9n_SGGS_NLS	A106V_D108N_D147Y_E155V
pNMG-239	pCMV_AAG*(wt)_XTEN_ecTadA_XTEN_Cas9n_SGGS_NLS	A106V_D108N_D147Y_E155V
pNMG-240	pCMV_ecTadA_XTEN_Cas9n_XTEN_EndoV*(D35A)_SGGS_NLS	A106V_D108N_D147Y_E155V
pNMG-241	pCMV_ecTadA_XTEN_Cas9n_XTEN_EndoV*(wt)_SGGS_NLS	A106V_D108N_D147Y_E155V
pNMG-242	pCMV_EndoV*(D35A)_XTEN_ecTadA_XTEN_Cas9n_SGGS_NLS	A106V_D108N_D147Y_E155V
pNMG-243	pCMV_EndoV*(wt)_XTEN_ecTadA_XTEN_Cas9n_SGGS_NLS	A106V_D108N_D147Y_E155V
pNMG-247	pCMV_ecTadA_XTEN_Cas9 (wild-type)_SGGS_NLS	wild-type
pNMG-248	pCMV_ecTadA_XTEN_Cas9 (wild-type)_SGGS_NLS	D108N_D147Y_E155V
pNMG-249	pCMV_ecTadA_XTEN_Cas9 (wild-type)_SGGS_NLS	A106V_D108N_D147Y_E155V
pNMG-250	pCMV_ecTadA_XTEN_Cas9 (wild-type)_SGGS_UGI_SGGS_NLS	D108N_D147Y_E155V
pNMG-251	pCMV_ecTadA_XTEN_Cas9 (wild-type)_SGGS_AAG*(E125Q)_SGGS_NLS	A106V_D108N_D147Y_E155V
pNMG-274	pCMV_ecTadA_SGGS_NLS (no Cas9 fusion)	wild-type
pNMG-275	pCMV_ecTadA_SGGS_NLS (no Cas9 fusion)	A106V_D108N_D147Y_E155V
pNMG-276	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2_ecTadA_XTEN_nCas9_SGGS_NLS	(wild-type) + (wild-type)

pNMG-277	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2_ecTadA_XTEN_nCas9_SGGS_NLS	(A106V_D108N_D147Y_E155V) + (A106V_D108N_D147Y_E155V)
pNMG-278	pCMV_ecTadA_XTEN_nCas9_SGGS_NLS	D108Q_D147Y_E155V
pNMG-279	pCMV_ecTadA_XTEN_nCas9_SGGS_NLS	D108M_D147Y_E155V
pNMG-280	pCMV_ecTadA_XTEN_nCas9_SGGS_NLS	D108L_D147Y_E155V
pNMG-281	pCMV_ecTadA_XTEN_nCas9_SGGS_NLS	D108K_D147Y_E155V
pNMG-282	pCMV_ecTadA_XTEN_nCas9_SGGS_NLS	D108I_D147Y_E155V
pNMG-283	pCMV_ecTadA_XTEN_nCas9_SGGS_NLS	D108F_D147Y_E155V
pNMG-284	pCMV_ecTadA_LONGER LINKER (92 a.a.)_ecTadA_XTEN_nCas9_SGGS_NLS	(wild-type) + (A106V_D108N_D147Y_E155V)
pNMG-285	pCMV_ecTadA_LONGER LINKER (92 a.a.)_ecTadA_XTEN_nCas9_SGGS_NLS	(A106V_D108N_D147Y_E155V) + (A106V_D108N_D147Y)
pNMG-285b	pCMV_ecTadA_LONGER LINKER (92 a.a.)_ecTadA_XTEN_nCas9_SGGS_NLS	(A106V_D108N_D147Y_E155V) + (A106V_D108N_D147Y_E155V)
pNMG-286	pCMV_ecTadA_XTEN_nCas9_SGGS_NLS	A106V_D108M_D147Y_E155V
pNMG-287	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2_ecTadA_XTEN-nCas9 (*S. aureus*)_SGGS_NLS	(A106V_D108N_D147Y_E155V) + (A106V_D108N_D147Y_E155V)
pNMG-289	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2_ecTadA_XTEN_nCas9_SGGS_UGI_NLS	(A106V_D108N_D147Y_E155V) + (A106V_D108N_D147Y_E155V)
pNMG-290	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2_ecTadA_(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_UGI_NLS	(A106V_D108N_D147Y_E155V) + (A106V_D108N_D147Y_E155V)
pNMG-293	pCMV_ecTadA_XTEN_Cas9n_SGGS_NLS	E59A_A106V_D108N_D147Y_E155V
pNMG-294	pCMV_ecTadA_XTEN_Cas9n_SGGS_NLS	E59A
pNMG-295	pCMV_ecTadA_SGGS_NLS (no Cas9 fusion)	E59A
pNMG-296	pCMV_ecTadA_SGGS_NLS (no Cas9 fusion)	E59A cat dead_A106V_D108N_D147Y_E155V
pNMG-297	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2_ecTadA_XTEN_nCas9_SGGS_NLS	(A106V_D108N_D147Y_E155V) + (wild-type)
pNMG-298	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2_ecTadA_XTEN_nCas9_SGGS_NLS	(D108M_D147Y_E155V) + (D108M_D147Y_E155V)
pNMG-320	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2_ecTadA_XTEN_nCas9_SGGS_NLS	(wild-type) + (A106V_D108N_D147Y_E155V)
pNMG-321	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2_ecTadA_XTEN_nCas9_SGGS_NLS	(E59A_A106V_D108N_D147Y_E155V) + (A106V_D108N_D147Y_E155V)

pNMG-322	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2_ecTadA_XTEN_nCas9_SGGS_NLS	(A106V_D108N_D147Y_E155V) + (E59A_A106V_D108N_D147Y_E155V)
pNMG-335	pCMV_TadA3p-XTEN-TadA2p-XTEN-nCas9-NLS	wild-type
pNMG-336	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_UGI_SGGS_NLS	L84F_A106V_D108N_H123Y_D147Y_E155V_I156Y
pNMG-337	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_UGI_SGGS_NLS	A106V_D108N_D147Y_E155V
pNMG-338	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_UGI_SGGS_NLS	L84F_A106V_D108N_H123Y_D147Y_E155V_I156F
pNMG-339	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2_ecTadA_(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_UGI_SGGS_NLS	(L84F_A106V_D108N_H123Y_D147Y_E155V_I156Y) + (L84F_A106V_D108N_H123Y_D147Y_E155V_I156Y)
pNMG-340	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2_ecTadA_(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_UGI_SGGS_NLS	(A106V_D108N_D147Y_E155V) + (A106V_D108N_D147Y_E155V)
pNMG-341	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2_ecTadA_(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_UGI_SGGS_NLS	(L84F_A106V_D108N_H123Y_D147Y_E155V_I156F) + (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)
pNMG-345	pCMV_*S. aureus*TadA-(SGGS)2-XTEN-(SGGS)2-*S. aureus*TadA-(SGGS)2-XTEN-(SGGS)2-nCas9_SGGS_NLS	wild-type
pNMG-346	pCMV_*S. aureus*TadA-(SGGS)2-XTEN-(SGGS)2-*S. aureus*TadA-(SGGS)2-XTEN-(SGGS)2-nCas9_SGGS_NLS	(D108N) + (D108N)
pNMG-347	pCMV_*S. aureus*TadA-(SGGS)2-XTEN-(SGGS)2-*S. aureus*TadA-(SGGS)2-XTEN-(SGGS)2-nCas9_SGGS_NLS	(D107A_D018N) + (D107A_D108N)
pNMG-348	pCMV_*S. aureus*TadA-(SGGS)2-XTEN-(SGGS)2-*S. aureus*TadA-(SGGS)2-XTEN-(SGGS)2-nCas9_SGGS_NLS	(G26P_D107A_D108N) + (G26P_D107A_D108N)
pNMG-349	pCMV_*S. aureus*TadA-(SGGS)2-XTEN-(SGGS)2-*S. aureus*TadA-(SGGS)2-XTEN-(SGGS)2-nCas9_sGGS_NLS	(G26P_D107A_D108N_S142A) +(G26P_D107A_D108N_S142A)
pNMG-350	pCMV_*S. aureus*TadA-(SGGS)2-XTEN-(SGGS)2-*S. aureus*TadA-(SGGS)2-XTEN-(SGGS)2-nCas9_SGGS_NLS	(D104A_D108N_S142A) + (D107A_D108N_S142A)
pNMG-351	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F)
pNMG-352	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(E25G_R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F)

pNMG-353	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(E25D_R26G_L84F_A106V_R107K_D108N_H123Y_A142N_A143G_D147Y_E155V_I156F)
pNMG-354	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(R26Q_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F)
pNMG-355	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(E25M_R26G_L84F_A106V_R107P_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F)
pNMG-356	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(R26C_L84F_A106V_R107H_D108N_H123Y_A142N_D147Y_E155V_I156F)
pNMG-357	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(L84F_A106V_D108N_H123Y_A142N_A143L_D147Y_E155V_I156F)
pNMG-358	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(R26G_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F)
pNMG-359	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(E25A_R26G_L84F_A106V_R107N_D108N_H123Y_A142N_A143E_D147Y_E155V_I156F)
pNMG-360	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F) +(R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F)
pNMG-361	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(E25G_R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F) X 2
pNMG-362	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(E25D_R26G_L84F_A106V_R107K_D108N_H123Y_A142N_A143G_D147Y_E155V_I156F) X 2
pNMG-363	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(R26Q_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F) X 2
pNMG-364	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(E25M_R26G_L84F_A106V_R107P_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F) X 2
pNMG-365	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(R26C_L84F_A106V_R107H_D108N_H123Y_A142N_D147Y_E155V_I156F) X 2
pNMG-366	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(L84F_A106V_D108N_H123Y_A142N_A143L_D147Y_E155V_I156F) X 2

pNMG-367	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(R26G_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F) X 2
pNMG-368	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(E25A_R26G_L84F_A106V_R107N_D108N_H123Y_A142N_A143E_D147Y_E155V_I156F) X 2
pNMG-369	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(L84F_A106V_D108N_H123Y_D147Y_E155V_I156Y) + (L84F_A106V_D108N_H123Y_D147Y_E155V_I156Y)
pNMG-370	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(A106V_D108N_D147Y_E155V) + (A106V_D108N_D147Y_E155V)
pNMG-371	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(L84F_A106V_D108N_H123Y_D147Y_E155V_I156F) + (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)
pNMG-372	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_NLS	A106V_D108N_A142N_D147Y_E155V
pNMG-373	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_NLS	R26G_A106V_D108N_A142N_D147Y_E155V
pNMG-374	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_NLS	E25D_R26G_A106V_R107K_D108N_A142N_A143G_D147Y_E155V
pNMG-375	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_NLS	R26G_A106V_D108N_R107H_A142N_A143D_D147Y_E155V
pNMG-376	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_NLS	E25D_R26G_A106V_D108N_A142N_D147Y_E155V
pNMG-377	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_NLS	A106V_R107K_D108N_A142N_D147Y_E155V
pNMG-378	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_NLS	A106V_D108N_A142N_A143G_D147Y_E155V
pNMG-379	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_NLS	A106V_D108N_A142N_A143L_D147Y_E155V
pNMG-382	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	A106V_D108N_A142N_D147Y_E155V X 2
pNMG-383	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	R26G_A106V_D108N_A142N_D147Y_E155V X 2
pNMG-384	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	E25D_R26G_A106V_R107K_D108N_A142N_A143G_D147Y_E155V X 2
pNMG-385	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	R26G_A106V_D108N_R107H_A142N_A143D_D147Y_E155V X 2

pNMG-386	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	E25D_R26G_A106V_D108N_A142N_D147Y_E155V X 2
pNMG-387	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	A106V_R107K_D108N_A142N_D147Y_E155V X 2
pNMG-388	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	A106V_D108N_A142N_A143G_D147Y_E155V X 2
pNMG-389	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	A106V_D108N_A142N_A143L_D147Y_E155V X 2
pNMG-391	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N
pNMG-392	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	N37T_P48T_M70L_L84F_A106V_D108N_H123Y_D147Y_I49V_E155V_I156F
pNMG-393	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	N37S_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K161T
pNMG-394	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	H36L_L84F_A106V_D108N_H123Y_D147Y_Q154H_E155V_I156F
pNMG-395	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	N72S_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F
pNMG-396	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	H36L_P48L_L84F_A106V_D108N_H123Y_E134G_D147Y_E155V_I156F
pNMG-397	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	H36L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N
pNMG-398	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	H36L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F
pNMG-399	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T
pNMG-400	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	N37S_R51H_D77G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F
pNMG-401	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	R51L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N
pNMG-402	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N) x 2

pNMG-403	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(N37T_P48T_M70L_L84F_A106V_D108N_H123Y_D147Y_I49V_E155V_I156F) x 2
pNMG-404	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(N37S_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K161T) x 2
pNMG-405	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(H36L_L84F_A106V_D108N_H123Y_D147Y_Q154H_E155V_I156F) x 2
pNMG-406	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(N72S_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F) x 2
pNMG-407	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(H36L_P48L_L84F_A106V_D108N_H123Y_E134G_D147Y_E155V_I156F) x 2
pNMG-408	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(H36L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N) x 2
pNMG-409	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(H36L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F) x 2
pNMG-410	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T) x 2
pNMG-411	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(N37S_R51H_D77G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F) x 2
pNMG-412	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(R51L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N) x 2
pNMG-440	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	D24G_Q71R_L84F_H96L_A106V_D108N_H123Y_D147Y_E155V_I156F_K160E
pNMG-441	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	H36L_G67V_L84F_A106V_D108N_H123Y_S146T_D147Y_E155V_I156F
pNMG-442	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	Q71L_L84F_A106V_D108N_H123Y_L137M_A143E_D147Y_E155V_I156F
pNMG-443	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	E25G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L
pNMG-444	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	L84F_A91T_F104I_A106V_D108N_H123Y_D147Y_E155V_I156F

pNMG-445	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	N72D_L84F_A106V_D108N_H123Y_G125A_D147Y_E155V_I156F
pNMG-446	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	P48S_L84F_S97C_A106V_D108N_H123Y_D147Y_E155V_I156F
pNMG-447	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	W23G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F
pNMG-448	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	D24G_P48L_Q71R_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L
pNMG-449	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(D24G_Q71R_L84F_H96L_A106V_D108N_H123Y_D147Y_E155V_I156F_K160E) x 2
pNMG-450	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(H36L_G67V_L84F_A106V_D108N_H123Y_S146T_D147Y_E155V_I156F) x 2
pNMG-451	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(Q71L_L84F_A106V_D108N_H123Y_L137M_A143E_D147Y_E155V_I156F) x 2
pNMG-452	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(E25G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L) x 2
pNMG-453	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(L84F_A91T_F104I_A106V_D108N_H123Y_D147Y_E155V_I156F) x 2
pNMG-454	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(N72D_L84F_A106V_D108N_H123Y_G125A_D147Y_E155V_I156F) x 2
pNMG-455	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(P48S_L84F_S97C_A106V_D108N_H123Y_D147Y_E155V_I156F) x 2
pNMG-456	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(W23G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F) x 2
pNMG-457	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(D24G_P48L_Q71R_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L) x 2
pNMG-473	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F
pNMG-474	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F) x 2

pNMG-475	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wild-typet) + (A106V_D108N_D147Y_E155V)
pNMG-476	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wild-type) + (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)
pNMG-477	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wild-type) + (H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N)
pNMG-478	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wild-type) + (N37S_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K161T)
pNMG-479	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wild-type) + (L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T)
pNMG-480	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_NLS	wild-type
pNMG-481	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_NLS	A106V_D108N
pNMG-482	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	wild-type + wild-type
pNMG-483	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(A106V_D108N) x2
pNMG-484	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wild-type) + (A106V_D108N)
pNMG-485	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	H36L_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F _K157N
pNMG-486	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	N37S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_K161T
pNMG-487	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	L84F_A106V_D108N_D147Y_E155V_I156F
pNMG-488	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K161T
pNMG-489	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K161T
pNMG-490	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E_K161T
pNMG-491	pCMV_ecTadA_(SGGS)2-XTEN-(SGGS)2_Cas9n_SGGS_UGI_SGGS_NLS	L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E
pNMG-492	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt) + (L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F)
pNMG-493	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt) + (D24G_Q71R_L84F_H96L_A106V_D108N_H123Y_D147Y_E155V_I156F_K160E)
pNMG-494	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt) + (H36L_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F _K157N)

pNMG-495	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt) + (N37S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_K161T)
pNMG-496	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt) +(L84F_A106V_D108N_D147Y_E155V_I156F)
pNMG-497	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt) + (R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K161T)
pNMG-498	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt) + (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K161T)
pNMG-499	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt) + (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E_K161T)
pNMG-500	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt) + (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E)
pNMG-513	pCMV_ecTadA-92 a.a.-ecTadA-32 a.a._nCas9_SGGS_NLS	(wt) + (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)
pNMG-514	pCMV_ecTadA-92 a.a.-ecTadA-32 a.a._nCas9_SGGS_NLS	(L84F_A106V_D108N_H123Y_D147Y_E155V_I156F) + (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)
pNMG-515	pCMV_ecTadA-92 a.a.-ecTadA- 64 a.a._nCas9_SGGS_NLS	(wt) + (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)
pNMG-516	pCMV_ecTadA-92 a.a.-ecTadA- 64 a.a._nCas9_SGGS_NLS	(L84F_A106V_D108N_H123Y_D147Y_E155V_I156F) + (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)
pNMG-517	pCMV_ecTadA- 32 a.a.-ecTadA- 64 a.a._nCas9_SGGS_NLS	(wt) + (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)
pNMG-518	pCMV_ecTadA- 32 a.a.-ecTadA- 64 a.a._nCas9_SGGS_NLS	(L84F_A106V_D108N_H123Y_D147Y_E155V_I156F) + (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)
pNMG-519	pCMV_ecTadA- 32 a.a.-_nCas9_SGGS_NLS	R74Q
pNMG-520	pCMV_ecTadA- 32 a.a.-_nCas9_SGGS_NLS	R74Q L84F_A106V_D108N_H123Y_D147Y_E155V_I156F
pNMG-521	pCMV_ecTadA- 32 a.a.-_nCas9_SGGS_NLS	R74A_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F
pNMG-522	pCMV_ecTadA- 32 a.a.-_nCas9_SGGS_NLS	R98Q
pNMG-523	pCMV_ecTadA- 32 a.a.-_nCas9_SGGS_NLS	R129Q
pNMG-524	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt + R74Q) + (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)
pNMG-525	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt + R74Q) + (R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)

pNMG-526	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(R74A_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F) + (R74A_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)
pNMG-527	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt + R98Q) + (L84F_R98Q_A106V_D108N_H123Y_D147Y_E155V_I156F)
pNMG-528	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt + R129Q) + (L84F_A106V_D108N_H123Y_R129Q_D147Y_E155V_I156F)
pNMG-529	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(L84F_A106V_D108N_H123Y_D147Y_E155V_I156F) + (H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N)
pNMG-530	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N) + (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)
pNMG-543	pCMV_ ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(P48S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F)
pNMG-544	pCMV_ ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(P48T_I49V_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_L157N)
pNMG-545	pCMV_ ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	P48S_A142N
pNMG-546	pCMV_ ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	P48T_I49V_A142N
pNMG-547	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt) + (P48S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F)
pNMG-548	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(P48S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F) + (P48S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F))
pNMG-549	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(P48S_A142N) + (P48S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F))
pNMG-550	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(P48S_A142N) + (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)
pNMG-551	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt) + (P48T_I49V_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_L157N)
pNMG-552	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(P48T_I49V_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_L157N) + (P48T_I49V_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_L157N)
pNMG-553	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(P48T_I49V_A142N) + (P48T_I49V_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_L157N)
pNMG-554	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(P48T_I49V_A142N) + (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)
pNMG-555	pCMV_ecTadA-24 a.a. linker-ecTadA-24 a.a. linker_nCas9_SGGS_NLS	(wt) + (H36L_R51L_L84F_A106V_D108N_H123Y_S146

		C_D147Y_E155V_I156F _K157N)
pNMG-556	pCMV_ecTadA-24 a.a. linker-ecTadA- 32 a.a. linker_nCas9_SGGS_NLS	(wt) + (H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N)
pNMG-557	pCMV_ecTadA-24 a.a. linker-ecTadA- 40 a.a. linker_nCas9_SGGS_NLS	(wt) + (H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N)
pNMG-558	pCMV_ecTadA- 32 a.a. linker-ecTadA- 24 a.a. linker_nCas9_SGGS_NLS	(wt) + (H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N)
pNMG-559	pCMV_ecTadA- 32 a.a. linker-ecTadA- 40 a.a. linker_nCas9_SGGS_NLS	(wt) + (H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N)
pNMG-560	pCMV_ecTadA- 40 a.a. linker-ecTadA- 24 a.a. linker_nCas9_SGGS_NLS	(wt) + (H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N)
pNMG-561	pCMV_ecTadA- 40 a.a. linker-ecTadA- 32 a.a. linker_nCas9_SGGS_NLS	(wt) + (H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N)
pNMG-562	pCMV_ecTadA- 40 a.a. linker-ecTadA- 40 a.a. linker_nCas9_SGGS_NLS	(wt) + (H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N)
pNMG-563	pCMV_ecTadA- 24 a.a. linker_nCas9_SGGS_NLS	wild-type
pNMG-564	pCMV_ecTadA- 24 a.a. linker_nCas9_SGGS_NLS	(H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N)
pNMG-565	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_XTEN_MBD4_SGGS_NLS	(wt) + (H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N)
pNMG-566	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_XTEN_TDG_SGGS_NLS	(wt) + (H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N)
pNMG-572	pCMV_ecTadA- 32 a.a.-_nCas9_SGGS_NLS	(H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N)
pNMG-573	pCMV_ecTadA- 32 a.a.-_nCas9_SGGS_NLS	(H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F _K157N)
pNMG-574	pCMV_ecTadA- 32 a.a.-_nCas9_SGGS_NLS	(H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N)
pNMG-575	pCMV_ecTadA- 32 a.a.-_nCas9_SGGS_NLS	(H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F _K157N)
pNMG-576	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt) + (H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N)
pNMG-577	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt) + (H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F _K157N)
pNMG-578	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt) + (H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N)

pNMG-579	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt) + (H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N)
pNMG-580	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N) + (H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)
pNMG-581	pCMV_ecTadA- 32 a.a.-_nCas9_SGGS_NLS	(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)
pNMG-583	pCMV_ecTadA- 32 a.a.-_nCas9_SGGS_NLS	(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N)
pNMG-586	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt) + (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)
pNMG-588	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt) + (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F_K157N)
pNMG-603	pCMV_ecTadA- 32 a.a.-_nCas9_SGGS_NLS	(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)
pNMG-604	pCMV_ecTadA- 32 a.a.-_nCas9_SGGS_NLS	(W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)
pNMG-605	pCMV_ecTadA- 32 a.a.-_nCas9_SGGS_NLS	(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T)
pNMG-606	pCMV_ecTadA- 32 a.a.-_nCas9_SGGS_NLS	(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152H_E155V_I156F_K157N)
pNMG-607	pCMV_ecTadA- 32 a.a.-_nCas9_SGGS_NLS	(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N)
pNMG-608	pCMV_ecTadA- 32 a.a.-_nCas9_SGGS_NLS	(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N)
pNMG-609	pCMV_ecTadA- 32 a.a.-_nCas9_SGGS_NLS	(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_E155V_I156F_K157N)
pNMG-610	pCMV_ecTadA- 32 a.a.-_nCas9_SGGS_NLS	(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_R152P_E155V_I156F_K157N)
pNMG-611	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt) + (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)
pNMG-612	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt) + (W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)
pNMG-613	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt) +(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T)
pNMG-614	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt) + (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152H_E155V_I156F_K157N)
pNMG-615	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt) + (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N)
pNMG-616	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt) + (W23L_H36L_P48A_R51L_L84F_A106V_D108N

		_H123Y_S146C_D147Y_R152P_E155V_I156F _K157N)
pNMG-617	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt) + (W23L_H36L_P48A_R51L_L84F_A106V_D108N _H123Y_A142A_S146C_D147Y_E155V_I156F _K157N)
pNMG-618	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt) + (W23L_H36L_P48A_R51L_L84F_A106V_D108N _H123Y_A142A_S146C_D147Y_R152P_E155V_ I156F _K157N)
pNMG-619	pCMV_ecTadA- 32 a.a.-_nCas9_SGGS_NLS	(W23R_H36L_P48A_R51L_L84F_A106V_D108N _H123Y_S146C_D147Y_R152P_E155V_I156F _K157N)
pNMG-620	pCMV_ecTadA-(SGGS)2-XTEN-(SGGS)2-ecTadA-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS	(wt) + (W23R_H36L_P48A_R51L_L84F_A106V_D108N _H123Y_S146C_D147Y_R152P_E155V_I156F _K157N)
pNMG-621	pCMV_ecTadA- 32 a.a. linker-ecTadA- 24 a.a. linker_nCas9_SGGS_NLS	(wt) + (H36L_P48A_R51L_L84F_A106V_D108N_H123Y _S146C_D147Y_R152P_E155V_I156F _K157N)
pNMG-622	pCMV_ecTadA- 32 a.a. linker-ecTadA- 24 a.a. linker_nCas9_SGGS_NLS	(wt) + (H36L_P48A_R51L_L84F_A106V_D108N_H123Y _A142N_S146C_D147Y_R152P_E155V_I156F _K157N)
pNMG-623	pCMV_ecTadA- 32 a.a. linker-ecTadA- 24 a.a. linker_nCas9_SGGS_NLS	(wt) + (W23L_H36L_P48A_R51L_L84F_A106V_D108N _H123Y_S146C_D147Y_R152P_E155V_I156F _K157N)
pNMG-624	pCMV_ecTadA- 32 a.a. linker-ecTadA- 24 a.a. linker_nCas9_SGGS_NLS	(wt) + (W23R_H36L_P48A_R51L_L84F_A106V_D108N _H123Y_S146C_D147Y_R152P_E155V_I156F _K157N)

[0056] In some embodiments, the A to G nucleobase editor comprises a dCas9 or nCas9 fused to an adenosine deaminase. Such nucleobase editors are described in US provisional application, U.S.S.N. 62/370,684, filed August 3, 2016; US provisional application, U.S.S.N. 62/370,684, filed February 3, 2017, US provisional application, U.S.S.N. 62/473,714, filed March 20, 2017, and PCT Application PCT/US2017/045381, filed August 3, 2017; each of which is incorporated herein by reference.

[0057] In some embodiments, an A to G nucleobase editor comprises the structure of NH2-[second adenosine deaminase]-[first adenosine deaminase]-[dCas9]-COOH. In some embodiments, the second adenosine deaminase is a wile-type ecTadA (SEQ ID NO: 314). In some embodiments, the a linker is used between each domain. In some embodiments, the linker is 32 amino acids long and comprises the amino acid sequence of SGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID NO: 384).

[0058] In some embodiments, the adenosine deaminase comprises one or more of a W23X, H36X, N37X, P48X, I49X, R51X, N72X, L84X, S97X, A106X, D108X, H123X, G125X, A142X, S146X, D147X, R152X, E155X, I156X, K157X, and/or K161X mutation in SEQ ID NO: 314, or one or more corresponding mutations in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of W23L, W23R, H36L, P48S, P48A, R51L, L84F, A106V, D108N, H123Y, A142N, S146C, D147Y, R152P, E155V, I156F, and/or K157N mutation in SEQ ID NO: 314, or one or more corresponding mutations in another adenosine deaminase.

[0059] In some embodiments, the adenosine deaminase comprises or consists of one, two, three, four, five, six, seven, eight, nine, ten, eleven, or twelve mutations selected from H36X, P48X, R51X, L84X, A106X, D108X, H123X, S146X, D147X, E155X, I156X, and K157X in SEQ ID NO: 314, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises or consists of one, two, three, four, five, six, seven, eight, nine, ten, eleven, or twelve mutations selected from H36L, P48S, R51L, L84F, A106V, D108N, H123Y, S146C, D147Y, E155V, I156F, and K157N in SEQ ID NO: 314, or a corresponding mutation or mutations in another adenosine deaminase. In some embodiments, the adenosine deaminse comprises or consists of a H36L, P48S, R51L, L84F, A106V, D108N, H123Y, S146C, D147Y, E155V, I156F, and K157N mutation in SEQ ID NO: 314, or corresponding mutations in another adenosine deaminase.

[0060] In some embodiments, the adenosine deaminase comprises or consists of one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, or thirteen mutations selected from H36X, P48X, R51X, L84X, A106X, D108X, H123X, A142X, S146X, D147X, E155X, I156X, and K157X in SEQ ID NO: 314, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises or consists of one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, or thirteen mutations selected from H36L, P48S, R51L, L84F, A106V, D108N, H123Y, A142N, S146C, D147Y, E155V, I156F, and K157N in SEQ ID NO: 314, or a corresponding mutation or mutations in another adenosine deaminase. In some embodiments, the adenosine

deaminase comprises or consists of a H36L, P48S, R51L, L84F, A106V, D108N, H123Y, A142N, S146C, D147Y, E155V, I156F, and K157N mutation in SEQ ID NO: 314, or corresponding mutations in another adenosine deaminase.

[0061] In some embodiments, the adenosine deaminase comprises or consists of one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, or fourteen mutations selected from W23X, H36X, P48X, R51X, L84X, A106X, D108X, H123X, A142X, S146X, D147X, E155X, I156X, and K157X in SEQ ID NO: 314, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises or consists of one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, or fourteen mutations selected from W23L, H36L, P48A, R51L, L84F, A106V, D108N, H123Y, A142N, S146C, D147Y, E155V, I156F, and K157N in SEQ ID NO: 314, or a corresponding mutation or mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises or consists of a W23L, H36L, P48A, R51L, L84F, A106V, D108N, H123Y, A142N, S146C, D147Y, E155V, I156F, and K157N mutation in SEQ ID NO: 314, or corresponding mutations in another adenosine deaminase.

[0062] In some embodiments, the adenosine deaminase comprises or consists of one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, or fifteen mutations selected from W23X, H36X, P48X, R51X, L84X, A106X, D108X, H123X, A142X, S146X, D147X, R152X, E155X, I156X, and K157X in SEQ ID NO: 314, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises or consists of one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, or fifteen mutations selected from W23L, H36L, P48A, R51L, L84F, A106V, D108N, H123Y, A142N, S146C, D147Y, R152P, E155V, I156F, and K157N in SEQ ID NO: 314, or a corresponding mutation or mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises or consists of a W23L, H36L, P48A, R51L, L84F, A106V, D108N, H123Y, A142N, S146C, D147Y, R152P, E155V, I156F, and K157N mutation in SEQ ID NO: 314, or corresponding mutations in another adenosine deaminase.

[0063] In some embodiments, the adenosine deaminase comprises or consists of one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, or fourteen mutations selected from W23X, H36X, P48X, R51X, L84X, A106X, D108X, H123X, S146X, D147X, R152X, E155X, I156X, and K157X in SEQ ID NO: 314, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises or consists of one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, or fourteen mutations selected from W23R, H36L, P48A, R51L, L84F, A106V, D108N, H123Y, S146C, D147Y, R152P, E155V, I156F, and K157N in SEQ ID NO: 314, or a corresponding mutation or mutations in another adenosine deaminase. In some embodiments, the adenosine deaminse comprises or consists of a W23R, H36L, P48A, R51L, L84F, A106V, D108N, H123Y, S146C, D147Y, R152P, E155V, I156F, and K157N mutation in SEQ ID NO: 314, or corresponding mutations in another adenosine deaminase.

[0064] In some embodiments, a nucleobase editor converts a C to G. Such nucleobase editors are described in US provisional application, U.S.S.N. 62/470, 175, filed March 10, 2017, US provisional application, U.S.S.N. 62/470, 175, filed March 10, 2017incorporated herein by reference.

[0065] Non-limiting, exemplary types of nucleobase editors (including C to T, A to G, and C to G nucleobase editors) and their respective sequences are provided in Example 1. In some embodiments, the nucleobase editor is a variant of the nucleobase editors described herein. For example, in some embodiments, the nucleobase editor is at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75% at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a nucleobase editor described herein (exemplary sequences are provided in Example 1). In some embodiments, the nucleobase editor comprises an amino acid sequence that is shorter or longer in length (*e.g.*, by no more than 30%, no more than 25%, no more than 20%, no more than 15%, no more than 10%, no more than 5%, no more than 1% longer or shorter) than any of the nucleobase editors provided herein. In some embodiments, the nucleobase editor comprises an amino acid sequence that is shorter or longer in length (*e.g.*, by no more than 500 amino acids, no more than 450 amino acids, no more than 400 amino acids, no more than 350 amino acids, no more than 300 amino acids, no more than 250 amino acids, no more than 200 amino acids, no

more than 200 amino acids, no more than 150 amino acids, no more than 100 amino acids, no more than 50 amino acids, no more than 10 amino acids, no more than 5 amino acids longer or shorter) than any of the nucleobase editors provided herein.

[0066] A "deaminase" refers to an enzyme that catalyzes the removal of an amine group from a molecule, or deamination, for example, through hydrolysis. In some embodiments, the deaminase is a cytidine deaminase, catalyzing the deamination of cytidine (C) to uridine (U), deoxycytidine (dC) to deoxyuridine (dU), or 5-methyl-cytidine to thymidine (T, 5-methyl-U), respectively. Subsequent DNA repair mechanisms ensure that a dU is replaced by T, as described in Komor *et al.* (*Nature*, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage, 533, 420–424 (2016), which is incorporated herein by reference). In some embodiments, the deaminase is a cytosine deaminase, catalyzing and promoting the conversion of cytosine to uracil (*e.g.*, in RNA) or thymine (*e.g.*, in DNA). In some embodiments, the deaminase is an adenosine deaminase that converts an A to G. In some embodiments, the deaminase is a naturally-occurring deaminase from an organism, such as a human, chimpanzee, gorilla, monkey, cow, dog, rat, or mouse. In some embodiments, the deaminase is a variant of a naturally-occurring deaminase from an organism, and the variants do not occur in nature. For example, in some embodiments, the deaminase or deaminase domain is at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75% at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a naturally-occurring deaminase from an organism. In some embodiments, the deaminase comprises an amino acid sequence that is shorter or longer in length (*e.g.*, by no more than 30%, no more than 25%, no more than 20%, no more than 15%, no more than 10%, no more than 5%, no more than 1% longer or shorter) than any of the deaminases provided herein. In some embodiments, the deaminase comprises an amino acid sequence that is shorter or longer in length (*e.g.*, by no more than 100 amino acids, no more than 90 amino acids, no more than 80 amino acids, no more than 70 amino acids, no more than 60 amino acids, no more than 50 amino acids, no more than 40 amino acids, no more than 30 amino acids, no more than 20 amino acids, no more than 10 amino acids, no more than 5 amino acids, no more than 2 amino acids, longer or shorter) than any of the deaminases provided herein.

[0067] A "split nucleobase editor" refers to a nucleobase editor that is provided as an N-terminal portion (also referred to as a N-terminal half) and a C-terminal portion (also referred to as a C-

terminal half) encoded by two separate nucleic acids. The polypeptides corresponding to the N-terminal portion and the C-terminal portion of the nucleobase editor may be combined to form a complete nucleobase editor. In some embodiments, for a nucleobase editor that comprises a dCas9 or nCas9, the "split" is located in the dCas9 or nCas9 domain, at positions as described herein in the split Cas9. Accordingly, in some embodiments, the N-terminal portion of the nucleobase editor contains the N-terminal portion of the split Cas9, and the C-terminal portion of the nucleobase editor contains the C-terminal portion of the split Cas9. Similarly, intein-N or intein-C may be fused to the N-terminal portion or the C-terminal portion of the nucleobase editor, respectively, for the joining of the N- and C-terminal portions of the nucleobase editor to form a complete nucleobase editor.

[0068] Two proteins or protein domains are considered to be "fused" when a peptide bond is formed linking the two proteins or two protein domains. In some embodiments, a linker (*e.g.*, a peptide linker) is present between the two proteins or two protein domains. The term "linker," as used herein, refers to a chemical group or a molecule linking two molecules or moieties, *e.g.*, two domains of a fusion protein, such as, for example, a nuclease-inactive Cas9 domain and a nucleic acid editing domain (*e.g.*, a deaminase domain). Typically, the linker is positioned between, or flanked by, two groups, molecules, or other moieties and connected to each one via a covalent bond, thus connecting the two. In some embodiments, the linker is an amino acid or a plurality of amino acids (*e.g.*, a peptide or protein). In some embodiments, the linker is an organic molecule, group, polymer, or chemical moiety. In some embodiments, the linker is 5-100 amino acids in length, for example, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, or 150-200 amino acids in length. Longer or shorter linkeare also contemplated.

[0069] A "uracil glycosylase inhibitor (UGI)" refers to a protein that inhibits the activity of uracil-DNA glycosylase. Suitable UGI proteins for use in accordance with the present disclosure include, for example, those published in Wang *et al.*, *J. Biol. Chem.* 264:1163-1171(1989); Lundquist *et al.*, *J. Biol. Chem.* 272:21408-21419(1997); Ravishankar *et al.*, *Nucleic Acids Res.* 26:4880-4887(1998); and Putnam *et al.*, *J. Mol. Biol.* 287:331-346(1999), each of which is incorporated herein by reference. Non-limiting, exemplary proteins that may be used as a UGI of the present disclosure and their respective sequences are provided in Example 1. In some

embodiments, the UGI is a variant of a naturally-occurring deaminase from an organism, and the variants do not occur in nature. For example, in some embodiments, the UGI is at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75% at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a naturally-occurring UGI from an organism or any UGIs provided herein (*e.g.*, in Example 1). In some embodiments, the UGI comprises an amino acid sequence that is shorter or longer in length (*e.g.*, by no more than 30%, no more than 25%, no more than 20%, no more than 15%, no more than 10%, no more than 5%, no more than 1% longer or shorter) than any of the UGIs provided herein. In some embodiments, the UGI comprises an amino acid sequence that is shorter or longer in length (*e.g.*, by no more than 20 amino acids, no more than 15 amino acids, no more than 10 amino acids, no more than 5 amino acids, no more than 2 amino acids longer or shorter) than any of the UGIs provided herein.

[0070] A gRNA is a component of the CRISPR/Cas system. A "gRNA" (guide ribonucleic acid) herein refers to a fusion of a CRISPR-targeting RNA (crRNA) and a trans-activation crRNA (tracrRNA), providing both targeting specificity and scaffolding/binding ability for Cas9 nuclease. A "crRNA" is a bacterial RNA that confers target specificity and requires tracrRNA to bind to Cas9. A "tracrRNA" is a bacterial RNA that links the crRNA to the Cas9 nuclease and typically can bind any crRNA. The sequence specificity of a Cas DNA-binding protein is determined by gRNAs, which have nucleotide base-pairing complementarity to target DNA sequences. The native gRNA comprises a 20 nucleotide (nt) Specificity Determining Sequence (SDS), which specifies the DNA sequence to be targeted, and is immediately followed by a 80 nt scaffold sequence, which associates the gRNA with Cas9. In some embodiments, an SDS of the present disclosure has a length of 15 to 100 nucleotides, or more. For example, an SDS may have a length of 15 to 90, 15 to 85, 15 to 80, 15 to 75, 15 to 70, 15 to 65, 15 to 60, 15 to 55, 15 to 50, 15 to 45, 15 to 40, 15 to 35, 15 to 30, or 15 to 20 nucleotides. In some embodiments, the SDS is 20 nucleotides long. For example, the SDS may be 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 nucleotides long. At least a portion of the target DNA sequence is complementary to the SDS of the gRNA. For Cas9 to successfully bind to the DNA target sequence, a region of the target sequence is complementary to the SDS of the gRNA sequence and is immediately followed by the correct protospacer adjacent motif (PAM) sequence (*e.g.*, NGG for Cas9 and TTN, TTTN, or YTN for Cpf1). In some embodiments, an SDS is 100% complementary to its

target sequence. In some embodiments, the SDS sequence is less than 100% complementary to its target sequence and is, thus, considered to be partially complementary to its target sequence. For example, a targeting sequence may be 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, or 90% complementary to its target sequence. In some embodiments, the SDS of template DNA or target DNA may differ from a complementary region of a gRNA by 1, 2, 3, 4 or 5 nucleotides.

[0071] In addition to the SDS, the gRNA comprises a scaffold sequence (corresponding to the tracrRNA in the native CRISPR/Cas system) that is required for its association with Cas9 (referred to herein as the "gRNA handle"). In some embodiments, the gRNA comprises a structure 5′-[SDS] -[gRNA handle]-3′. In some embodiments, the scaffold sequence comprises the nucleotide sequence of 5′-guuuuagagcuagaaauagcaaguuaaaauaaaggcuaguc cguuaucaacuugaaaaaguggcaccgagucggugcuuuuu-3′ (SEQ ID NO: 358). Other non-limiting, suitable gRNA handle sequences that may be used in accordance with the present disclosure are listed in Table 1.

Table 1. Guide RNA Handle Sequences

Organism	**gRNA handle sequence**	**SEQ ID NO**
S. pyogenes	GUUUAAGAGCUAUGCUGGAAACAGCACGGUGAAA AAGUUCAACUAUUGCCUGAUCGGAAUAAAUUUG AACGAUACGACAGUCGGUGCUUUUUUU	359
S. pyogenes	GUUUAAGAGCUAGAAAUAGCAAGUUUAAAUAA GGCUAGUCCGUUAUCAACUUGAAAAAGUGGCAC CGAGUCGGUGCUUUUUU	360
S. thermophilus CRISPR1	GUUUUUGUACUCUCAAGAUUCAAUAAUCUUGCA GAAGCUACAAAGAUAAGGCUUCAUGCCGAAAUC AACACCCUGUCAUUUUAUGGCAGGGUGUUUU	361
S. thermophilus CRISPR3	GUUUUAGAGCUGUGUUGUUUGUUAAAACAACA CAGCGAGUUAAAAUAAGGCUUAGUCCGUACUCA ACUUGAAAAGGUGGCACCGAUUCGGUGUUUUU	362
C. jejuni	AAGAAAUUUAAAAAGGGACUAAAAUAAAGAGU UUGCGGGACUCUGCGGGGUUACAAUCCCCUAAA ACCGCUUUU	363
F. novicida	AUCUAAAAUUAUAAAUGUACCAAAUAAUUAAU GCUCUGUAAUCAUUUAAAAGUAUUUUGAACGG ACCUCUGUUUGACACGUCUGAAUAACUAAAA	364
S. thermophilus2	UGUAAGGGACGCCUUACACAGUUACUUAAAUCU UGCAGAAGCUACAAAGAUAAGGCUUCAUGCCGA	365

	AAUCAACACCCUGUCAUUUUAUGGCAGGGUGUU UUCGUUAUUU	
M. mobile	UGUAUUUCGAAAUACAGAUGUACAGUUAAGAA UACAUAAGAAUGAUACAUCACUAAAAAAAGGCU UUAUGCCGUAACUACUACUUAUUUUCAAAAUAA GUAGUUUUUUUU	366
L. innocua	AUUGUUAGUAUUCAAAAUAACAUAGCAAGUUA AAAUAAGGCUUUGUCCGUUAUCAACUUUUAAUU AAGUAGCGCUGUUUCGGCGCUUUUUUU	367
S. pyogenes	GUUGGAACCAUUCAAAACAGCAUAGCAAGUUAA AAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGU GGCACCGAGUCGGUGCUUUUUUU	368
S. mutans	GUUGGAAUCAUUCGAAACAACACAGCAAGUUAA AAUAAGGCAGUGAUUUUUAAUCCAGUCCGUACA CAACUUGAAAAAGUGCGCACCGAUUCGGUGCUU UUUUAUUU	369
S. thermophilus	UUGUGGUUUGAAACCAUUCGAAACAACACAGCG AGUUAAAAUAAGGCUUAGUCCGUACUCAACUUG AAAAGGUGGCACCGAUUCGGUGUUUUUUUU	370
N. meningitidis	ACAUAUUGUCGCACUGCGAAAUGAGAACCGUUG CUACAAUAAGGCCGUCUGAAAAGAUGUGCCGCA ACGCUCUGCCCCUUAAAGCUUCUGCUUUAAGGG GCA	371
P. multocida	GCAUAUUGUUGCACUGCGAAAUGAGAGACGUUG CUACAAUAAGGCUUCUGAAAAGAAUGACCGUAA CGCUCUGCCCCUUGUGAUUCUUAAUUGCAAGGG GCAUCGUUUUU	372

[0072] In some embodiments, the guide RNA is about 15-120 nucleotides long and comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence. In some embodiments, the guide RNA is 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, or 120 nucleotides long. In some embodiments, the guide RNA comprises a sequence of 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more contiguous nucleotides that is complementary to a target sequence. Sequence complementarity refers to distinct interactions between adenine and thymine (DNA) or uracil (RNA), and between guanine and cytosine.

[0073] A "protospacer adjacent motif" (PAM) is typically a sequence of nucleotides located adjacent to (*e.g.*, within 10, 9, 8, 7, 6, 5, 4, 3, 3, or 1 nucleotide(s) of a target sequence). A PAM sequence is "immediately adjacent to" a target sequence if the PAM sequence is contiguous with the target sequence (that is, if there are no nucleotides located between the PAM sequence and the target sequence). In some embodiments, a PAM sequence is a wild-type PAM sequence. Examples of PAM sequences include, without limitation, NGG, NGR, NNGRR(T/N), NNNNGATT, NNAGAAW, NGGAG, NAAAAC, AWG, and CC. In some embodiments, a PAM sequence is obtained from *Streptococcus pyogenes* (*e.g.*, NGG or NGR). In some embodiments, a PAM sequence is obtained from *Staphylococcus aureus* (*e.g.*, NNGRR(T/N)). In some embodiments, a PAM sequence is obtained from *Neisseria meningitidis* (*e.g.*, NNNNGATT). In some embodiments, a PAM sequence is obtained from *Streptococcus thermophilus* (*e.g.*, NNAGAAW or NGGAG). In some embodiments, a PAM sequence is obtained from *Treponema denticola* (*e.g.*, NAAAAC). In some embodiments, a PAM sequence is obtained from *Escherichia coli* (*e.g.*, AWG). In some embodiments, a PAM sequence is obtained from *Pseudomonas auruginosa* (*e.g.*, CC). Other PAM sequences are contemplated. A PAM sequence is typically located downstream (*i.e.*, 3′) from the target sequence, although in some embodiments a PAM sequence may be located upstream (*i.e.*, 5′) from the target sequence.

[0074] A "nuclear localization signal" or "NLS" refers to as an amino acid sequence that "tags" a protein for import into the cell nucleus by nuclear transport. Typically, this signal consists of one or more short sequences of positively charged lysines or arginines exposed on the protein surface. One or more NLS may be added to the N- or C-terminus of a protein, or internally (*e.g.*, between two protein domains). For example, one or more NLS may be added to the N- or C-terminus of a nucleobase editor, or between the Cas9 and the deaminase in a nucleobase editor. In some embodiments, 1, 2, 3, 4, 5, or more NLS may be added. Nuclear localization sequences are known in the art and would be apparent to the skilled artisan. For example, NLS sequences are described in Plank *et al.*, PCT/EP2000/011690, filed November 23, 2000, the contents of which are incorporated herein by reference for its disclosure of exemplary nuclear localization sequences. In some embodiments, a NLS comprises the amino acid sequence PKKKRKV (SEQ ID NO: 373) or MDSLLMNRRKFLYQFKNVRWAKGRRETYLC (SEQ ID NO: 374). In some embodiments, a linker is inserted between the Cas9 and the deaminase.

[0075] An NLS can be classified as monopartite or bipartite. A non-limiting example of a monopartite NLS is the sequence PKKKRKV (SEQ ID NO: 373) in the SV40 Large T-antigen. A bipartite NLS typically contains two clusters of basic amino acids, separated by a spacer of about 10 amino acids. One non-limiting example of a bipartite NLS is the NLS of nucleoplasmin, KRPAATKKAGQAKKKK (spacer underlined) (SEQ ID NO: 344). In some embodiments, the NLS used in accordance with the present disclosure is the NLS of nucleoplasmin comprising the amino acid sequence of KRPAATKKAGQAKKKK (SEQ ID NO: 344). Other bipartite NLSs that may be used in accordance with the present disclosure include, without limitation: SV40 bipartite NLS (KRTADGSEFESPKKKRKV (SEQ ID NO: 375), *e.g.*, as described in Hodel *et al.*, J Biol Chem. 2001 Jan 12;276(2):1317-25, incorporated herein by reference); Kanadaptin bipartite NLS (KKTELQTTNAENKTKKL (SEQ ID NO: 345), *e.g.*, as described in Hubner *et al.*, Biochem J. 2002 Jan 15;361(Pt 2):287-96, incorporated herein by reference); influenza A nucleoprotein bipartite NLS (KRGINDRNFWRGENGRKTR (SEQ ID NO: 346), *e.g.*, as described in Ketha *et al.*, BMC Cell Biology. 2008;9:22, incorporated herein by reference); and ZO-2 bipartite NLS (RKSGKIAAIVVKRPRK (SEQ ID NO: 347), *e.g.*, as described in Quiros *et al.*, Nusrat A, ed. Molecular Biology of the Cell. 2013;24(16):2528-2543, incorporated herein by reference).

[0076] The nucleotide sequence encoding an NLS is "operably linked" to the nucleotide sequence encoding a protein to which the NLS is fused (*e.g.*, a Cas9 or a nucleobase editor) when two coding sequences are "in-frame with each other" and are translated as a single polypeptide fusing two sequences.

[0077] Nucleic acids of the present disclosure may include one or more genetic elements. A "genetic element" refers to a particular nucleotide sequence that has a role in nucleic acid expression (*e.g.*, promoter, enhancer, terminator) or encodes a discrete product of an engineered nucleic acid (*e.g.*, a nucleotide sequence encoding a guide RNA, a protein and/or an RNA interference molecule).

[0078] A "promoter" refers to a control region of a nucleic acid sequence at which initiation and rate of transcription of the remainder of a nucleic acid sequence are controlled. A promoter may also contain sub-regions at which regulatory proteins and molecules may bind, such as RNA polymerase and other transcription factors. Promoters may be constitutive, inducible, activatable, repressible, tissue-specific, or any combination thereof. A promoter drives

expression or drives transcription of the nucleic acid sequence that it regulates. Herein, a promoter is considered to be "operably linked" when it is in a correct functional location and orientation in relation to a nucleic acid sequence it regulates to control ("drive") transcriptional initiation and/or expression of that sequence.

[0079] A promoter may be one naturally associated with a gene or sequence, as may be obtained by isolating the 5′ non-coding sequences located upstream of the coding segment of a given gene or sequence. Such a promoter is referred to as an "endogenous promoter." In some embodiments, a coding nucleic acid sequence may be positioned under the control of a recombinant or heterologous promoter, which refers to a promoter that is not normally associated with the encoded sequence in its natural environment. Such promoters may include promoters of other genes; promoters isolated from any other cell; and synthetic promoters or enhancers that are not "naturally occurring" such as, for example, those that contain different elements of different transcriptional regulatory regions and/or mutations that alter expression through methods of genetic engineering that are known in the art. In addition to producing nucleic acid sequences of promoters and enhancers synthetically, sequences may be produced using recombinant cloning and/or nucleic acid amplification technology, including polymerase chain reaction (PCR).

[0080] In some embodiments, promoters used in accordance with the present disclosure are "inducible promoters," which are promoters that are characterized by regulating (*e.g.*, initiating or activating) transcriptional activity when in the presence of, influenced by or contacted by an inducer signal. An inducer signal may be endogenous or a normally exogenous condition (*e.g.*, light), compound (*e.g.*, chemical or non-chemical compound) or protein that contacts an inducible promoter in such a way as to be active in regulating transcriptional activity from the inducible promoter. Thus, a "signal that regulates transcription" of a nucleic acid refers to an inducer signal that acts on an inducible promoter. A signal that regulates transcription may activate or inactivate transcription, depending on the regulatory system used. Activation of transcription may involve directly acting on a promoter to drive transcription or indirectly acting on a promoter by inactivation a repressor that is preventing the promoter from driving transcription. Conversely, deactivation of transcription may involve directly acting on a promoter to prevent transcription or indirectly acting on a promoter by activating a repressor that then acts on the promoter.

[0081] A "transcriptional terminator" is a nucleic acid sequence that causes transcription to stop. A transcriptional terminator may be unidirectional or bidirectional. It is comprised of a DNA sequence involved in specific termination of an RNA transcript by an RNA polymerase. A transcriptional terminator sequence prevents transcriptional activation of downstream nucleic acid sequences by upstream promoters. A transcriptional terminator may be necessary *in vivo* to achieve desirable expression levels or to avoid transcription of certain sequences. A transcriptional terminator is considered to be "operably linked to" a nucleotide sequence when it is able to terminate the transcription of the sequence it is linked to.

[0082] The most commonly used type of terminator is a forward terminator. When placed downstream of a nucleic acid sequence that is usually transcribed, a forward transcriptional terminator will cause transcription to abort. In some embodiments, bidirectional transcriptional terminators are provided, which usually cause transcription to terminate on both the forward and reverse strand. In some embodiments, reverse transcriptional terminators are provided, which usually terminate transcription on the reverse strand only.

[0083] In prokaryotic systems, terminators usually fall into two categories (1) rho-independent terminators and (2) rho-dependent terminators. Rho-independent terminators are generally composed of palindromic sequence that forms a stem loop rich in G-C base pairs followed by several T bases. Without wishing to be bound by theory, the conventional model of transcriptional termination is that the stem loop causes RNA polymerase to pause, and transcription of the poly-A tail causes the RNA:DNA duplex to unwind and dissociate from RNA polymerase.

[0084] In eukaryotic systems, the terminator region may comprise specific DNA sequences that permit site-specific cleavage of the new transcript so as to expose a polyadenylation site. This signals a specialized endogenous polymerase to add a stretch of about 200 A residues (polyA) to the 3' end of the transcript. RNA molecules modified with this polyA tail appear to more stable and are translated more efficiently. Thus, in some embodiments involving eukaryotes, a terminator may comprise a signal for the cleavage of the RNA. In some embodiments, the terminator signal promotes polyadenylation of the message. The terminator and/or polyadenylation site elements may serve to enhance output nucleic acid levels and/or to minimize read through between nucleic acids.

[0085] Terminators for use in accordance with the present disclosure include any terminator of transcription described herein or known to one of ordinary skill in the art. Examples of terminators include, without limitation, the termination sequences of genes such as, for example, the bovine growth hormone terminator, and viral termination sequences such as, for example, the SV40 terminator, spy, yejM, secG-leuU, thrLABC, rrnB T1, hisLGDCBHAFI, metZWV, rrnC, xapR, aspA and arcA terminator. In some embodiments, the termination signal may be a sequence that cannot be transcribed or translated, such as those resulting from a sequence truncation.

[0086] A "Woodchuck Hepatitis Virus (WHP) Posttranscriptional Regulatory Element (WPRE)" is a DNA sequence that, when transcribed creates a tertiary structure enhancing expression. Commonly used in molecular biology to increase expression of genes delivered by viral vectors. WPRE is a tripartite regulatory element with gamma, alpha, and beta components. The full WPRE sequence is 609bp long:

GCTTATCGATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCT
TAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCAT
GCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTC
TCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTT
TGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGG
GACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCC
CGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGG
AAATCATCGTCCTTTCCTTGGCTGCTCGCCTATGTTGCCACCTGGATTCTGCGCGGGA
CGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCT
GCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGAT
CTCCCTTTGGGCCGCCTCCCCGCATCGATACCG (SEQ ID NO:376).

[0087] An "adeno-associated virus" or "AAV" is a virus which infects humans and some other primate species. The wild-type AAV genome is a single-stranded deoxyribonucleic acid (ssDNA), either positive- or negative-sensed. The genome comprises two inverted terminal repeats (ITRs), one at each end of the DNA strand, and two open reading frames (ORFs): rep and cap between the ITRs. The rep ORF comprises four overlapping genes encoding Rep proteins required for the AAV life cycle. The cap ORF comprises overlapping genes encoding capsid proteins: VP1, VP2 and VP3, which interact together to form the viral capsid. VP1, VP2 and VP3 are translated from one mRNA transcript, which can be spliced in two different manners: either a longer or shorter intron can be excised resulting in the formation of two isoforms of mRNAs: a ~2.3 kb- and a ~2.6 kb-long mRNA isoform. The capsid forms a supramolecular assembly of approximately 60 individual capsid protein subunits into a non-enveloped, T-1

icosahedral lattice capable of protecting the AAV genome. The mature capsid is composed of VP1, VP2, and VP3 (molecular masses of approximately 87, 73, and 62 kDa respectively) in a ratio of about 1:1:10.

[0088] rAAV particles may comprise a nucleic acid vector (*e.g.*, a recombinant genome), which may comprise at a minimum: (a) one or more heterologous nucleic acid regions comprising a sequence encoding a protein or polypeptide of interest (*e.g.*, a split Cas9 or split nucleobase) or an RNA of interest (*e.g.*, a gRNA), or one or more nucleic acid regions comprising a sequence encoding a Rep protein; and (b) one or more regions comprising inverted terminal repeat (ITR) sequences (*e.g.*, wild-type ITR sequences or engineered ITR sequences) flanking the one or more nucleic acid regions (*e.g.*, heterologous nucleic acid regions). In some embodiments, the nucleic acid vector is between 4 kb and 5 kb in size (*e.g.*, 4.2 to 4.7 kb in size). In some embodiments, the nucleic acid vector further comprises a region encoding a Rep protein. In some embodiments, the nucleic acid vector is circular. In some embodiments, the nucleic acid vector is single-stranded. In some embodiments, the nucleic acid vector is double-stranded. In some embodiments, a double-stranded nucleic acid vector may be, for example, a self-complimentary vector that contains a region of the nucleic acid vector that is complementary to another region of the nucleic acid vector, initiating the formation of the double-strandedness of the nucleic acid vector.

[0089] The terms "nucleic acid," and "polynucleotide," as used herein, refer to a compound comprising a nucleobase and an acidic moiety, *e.g.*, a nucleotide, or a polymer of nucleotides. Typically, polymeric nucleic acids, *e.g.*, nucleic acid molecules comprising three or more nucleotides are linear molecules, in which adjacent nucleotides are linked to each other via a phosphodiester linkage. In some embodiments, "nucleic acid" refers to individual nucleic acid residues (*e.g.* nucleotides and/or nucleosides). In some embodiments, "nucleic acid" refers to an oligonucleotide chain comprising three or more individual nucleotide residues. As used herein, the terms "oligonucleotide" and "polynucleotide" can be used interchangeably to refer to a polymer of nucleotides (*e.g.*, a string of at least three nucleotides). In some embodiments, "nucleic acid" encompasses RNA as well as single and/or double-stranded DNA. Nucleic acids may be naturally occurring, for example, in the context of a genome, a transcript, an mRNA, tRNA, rRNA, siRNA, snRNA, a plasmid, cosmid, chromosome, chromatid, or other naturally occurring nucleic acid molecule. On the other hand, a nucleic acid molecule may be a non-

naturally occurring molecule, *e.g.*, a recombinant DNA or RNA, an artificial chromosome, an engineered genome (*e.g.,* an engineered viral vector), an engineered vector, or fragment thereof, or a synthetic DNA, RNA, or DNA/RNA hybrid, optionally including non-naturally occurring nucleotides or nucleosides. Furthermore, the terms "nucleic acid," "DNA," "RNA," and/or similar terms include nucleic acid analogs, *e.g.,* analogs having other than a phosphodiester backbone. Nucleic acids can be purified from natural sources, produced using recombinant expression systems and optionally purified, chemically synthesized, *etc.* Where appropriate, *e.g.*, in the case of chemically synthesized molecules, nucleic acids can comprise nucleoside analogs such as analogs having chemically modified bases or sugars, and backbone modifications. A nucleic acid sequence is presented in the 5′ to 3′ direction unless otherwise indicated. In some embodiments, a nucleic acid is or comprises natural nucleosides (*e.g.* adenosine, thymidine, guanosine, cytidine, uridine, deoxyadenosine, deoxythymidine, deoxyguanosine, and deoxycytidine); nucleoside analogs (*e.g.*, 2-aminoadenosine, 2-thiothymidine, inosine, pyrrolo-pyrimidine, 3-methyl adenosine, 5-methylcytidine, 2-aminoadenosine, C5-bromouridine, C5-fluorouridine, C5-iodouridine, C5-propynyl-uridine, C5-propynyl-cytidine, C5-methylcytidine, 2-aminoadenosine, 7-deazaadenosine, 7-deazaguanosine, 8-oxoadenosine, 8-oxoguanosine, O(6)-methylguanine, and 2-thiocytidine); chemically modified bases; biologically modified bases (*e.g.*, methylated bases); intercalated bases; modified sugars (*e.g.*, 2′-fluororibose, ribose, 2′-deoxyribose, arabinose, and hexose); and/or modified phosphate groups (*e.g.*, phosphorothioates and 5′-*N*-phosphoramidite linkages).

[0090] The terms "protein," "peptide," and "polypeptide" are used interchangeably herein, and refer to a polymer of amino acid residues linked together by peptide (amide) bonds. The terms refer to a protein, peptide, or polypeptide of any size, structure, or function. Typically, a protein, peptide, or polypeptide will be at least three amino acids long. A protein, peptide, or polypeptide may refer to an individual protein or a collection of proteins. One or more of the amino acids in a protein, peptide, or polypeptide may be modified, for example, by the addition of a chemical entity such as a carbohydrate group, a hydroxyl group, a phosphate group, a farnesyl group, an isofarnesyl group, a fatty acid group, a linker for conjugation, functionalization, or other modification, *etc.* A protein, peptide, or polypeptide may also be a single molecule or may be a multi-molecular complex. A protein, peptide, or polypeptide may be just a fragment of a naturally occurring protein or peptide. A protein, peptide, or polypeptide may be naturally

occurring, recombinant, or synthetic, or any combination thereof. The term "fusion protein" as used herein refers to a hybrid polypeptide which comprises protein domains from at least two different proteins. One protein may be located at the amino-terminal (N-terminal) portion of the fusion protein or at the carboxy-terminal (C-terminal) protein thus forming an "amino-terminal fusion protein" or a "carboxy-terminal fusion protein," respectively. A protein may comprise different domains, for example, a nucleic acid binding domain (*e.g.*, the gRNA binding domain of Cas9 that directs the binding of the protein to a target site) and a nucleic acid cleavage domain or a catalytic domain of a nucleic-acid editing protein. In some embodiments, a protein is in a complex with, or is in association with, a nucleic acid, *e.g.*, RNA or DNA. Any of the proteins provided herein may be produced by any method known in the art. For example, the proteins provided herein may be produced via recombinant protein expression and purification, which is especially suited for fusion proteins comprising a peptide linker. Methods for recombinant protein expression and purification are well known, and include those described by Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4^{th} ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)), which are incorporated herein by reference.

[0091] The term "subject," as used herein, refers to an individual organism, for example, an individual mammal. In some embodiments, the subject is a human. In some embodiments, the subject is a non-human mammal. In some embodiments, the subject is a non-human primate. In some embodiments, the subject is a rodent (*e.g.*, mouse, rat). In some embodiments, the subject is a domesticated animal. In some embodiments, the subject is a sheep, a goat, a cow, a cat, or a dog. In some embodiments, the subject is a research animal. In some embodiments, the subject is genetically engineered, *e.g.*, a genetically engineered non-human subject. The subject may be of either sex and at any stage of development.

[0092] The term "recombinant" as used herein in the context of proteins or nucleic acids refers to proteins or nucleic acids that do not occur in nature, but are the product of human engineering. For example, in some embodiments, a recombinant protein or nucleic acid molecule comprises an amino acid or nucleotide sequence that comprises at least one, at least two, at least three, at least four, at least five, at least six, or at least seven mutations as compared to any naturally occurring sequence. The fusion proteins (*e.g.*, base editors) described herein are made recombinantly. Recombinant technology is familiar to those skilled in the art.

[0093] The term "pharmaceutically-acceptable carrier" means a pharmaceutically-acceptable material, composition or vehicle, such as a liquid or solid filler, diluent, excipient, manufacturing aid (*e.g.*, lubricant, talc magnesium, calcium or zinc stearate, or steric acid), or solvent encapsulating material, involved in carrying or transporting the compound from one site (*e.g.*, the delivery site) of the body, to another site (*e.g.*, organ, tissue or portion of the body). A pharmaceutically acceptable carrier is "acceptable" in the sense of being compatible with the other ingredients of the formulation and not injurious to the tissue of the subject (*e.g.*, physiologically compatible, sterile, physiologic pH, *etc.*).

[0094] "A therapeutically effective amount" as used herein refers to the amount of each therapeutic agent (*e.g.*, nucleobase editor, rAAV) described in the present disclosure required to confer therapeutic effect on the subject, either alone or in combination with one or more other therapeutic agents. Effective amounts vary, as recognized by those skilled in the art, depending on the particular condition being treated, the severity of the condition, the individual subject parameters including age, physical condition, size, gender, and weight, the duration of the treatment, the nature of concurrent therapy (if any), the specific route of administration and like factors within the knowledge and expertise of the health practitioner. These factors are well known to those of ordinary skill in the art and can be addressed with no more than routine experimentation. It is generally preferred that a maximum dose of the individual components or combinations thereof be used, that is, the highest safe dose according to sound medical judgment. It will be understood by those of ordinary skill in the art, however, that a subject may insist upon a lower dose or tolerable dose for medical reasons, psychological reasons or for virtually any other reasons. Empirical considerations, such as the half-life, generally will contribute to the determination of the dosage. For example, therapeutic agents that are compatible with the human immune system, such as polypeptides comprising regions from humanized antibodies or fully human antibodies, may be used to prolong half-life of the polypeptide and to prevent the polypeptide being attacked by the host's immune system.

[0095] "A subject in need thereof" refers to an individual who has a disease, a sign and/or symptom of a disease, or a predisposition toward a disease, with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve, or affect the disease, the symptom of the disease, or the predisposition toward the disease. In some embodiments, the subject is a mammal. In some embodiments, the subject is a non-human primate. In some embodiments, the

subject is human. In some embodiments, the mammal is a rodent. In some embodiments, the rodent is a mouse. In some embodiments, the rodent is a rat. In some embodiments, the mammal is a companion animal. A "companion animal" refers to pets and other domestic animals. Non-limiting examples of companion animals include dogs and cats; livestock, such as horses, cattle, pigs, sheep, goats, and chickens; and other animals, such as mice, rats, guinea pigs, and hamsters.

[0001] The terms "treatment," "treat," and "treating," refer to a clinical intervention aimed to reverse, alleviate, delay the onset of, or inhibit the progress of a disease or disorder, or one or more symptoms thereof, as described herein. As used herein, the terms "treatment," "treat," and "treating" refer to a clinical intervention aimed to reverse, alleviate, delay the onset of, or inhibit the progress of a disease or disorder, or one or more symptoms thereof, as described herein. In some embodiments, treatment may be administered after one or more symptoms have developed and/or after a disease has been diagnosed. In other embodiments, treatment may be administered in the absence of symptoms, *e.g.*, to prevent or delay onset of a symptom or inhibit onset or progression of a disease. For example, treatment may be administered to a susceptible individual prior to the onset of symptoms (*e.g.*, in light of a history of symptoms and/or in light of genetic or other susceptibility factors). Treatment may also be continued after symptoms have resolved, for example, to prevent or delay their recurrence.

DETAILED DESCRIPTION OF CERTAIN EMBODIMENTS

[0096] Provided herein, are compositions (*e.g.*, vectors, recombinant viruses) and kits comprising nucleic acids encoding split Cas9 proteins or nucleobase editors, and methods of delivering a nucleobase editor or a Cas9 protein into a cell using such nucleic acids. The N-terminal portion and C-terminal portion of a nucleobase editor or a Cas9 protein are encoded on separate nucleic acids and delivered into a cell, *e.g.*, a via recombinant adeno-associated virus (rAAV particles) delivery. The polypeptides corresponding to the N-terminal portion and C-terminal portions of the nucleobase editor or Cas9 protein may be joined to form a complete nucleobase editor or Cas9 protein, *e.g.*, via intein-mediated protein splicing.

[0097] Accordingly, some aspects of the present disclosure relate to compositions comprising (i) a first nucleotide sequence encoding an N-terminal portion of a Cas9 protein fused at its C-terminus to an intein-N; and (ii) a second nucleotide sequence encoding an intein-C fused to the

N-terminus of a C-terminal portion of the Cas9 protein, wherein the first nucleotide sequence or second nucleotide sequence is operably linked to a nucleotide sequence encoding at least one bipartite nuclear localization signal (NLS).

[0098] The Cas9 protein encoded by the first and second nucleotide sequence is herein referred as a "split Cas9." The Cas9 protein is known to have a N-terminal lobe and a C-terminal lobe linked by a disordered linker (*e.g.*, as described in Nishimasu *et al.*, Cell, Volume 156, Issue 5, pp. 935–949, 2014, incorporated herein by reference). In some embodiments, the N-terminal portion of the split Cas9 protein comprises the N-terminal lobe of a Cas9 protein. In some embodiments, the C-terminal portion of the split Cas9 comprises the C-terminal lobe of a Cas9 protein. In some embodiments, the N-terminal portion of the split Cas9 comprises a portion of any one of SEQ ID NO: 1-275 and 394-397 that corresponds to amino acids 1-(550-650) in SEQ ID NO: 1. "1-(550-650)" means starting from amino acid 1 and ending anywhere between amino acid 550-650 (inclusive). For example, the N-terminal portion of the split Cas9 may comprise a portion of any one of SEQ ID NOs: 1-275 and 394-397 that corresponds to amino acids 1-550, 1-551, 1-552, 1-553, 1-554, 1-555, 1-556, 1-557, 1-558, 1-559, 1-560, 1-561, 1-562, 1-563, 1-564, 1-565, 1-566, 1-567, 1-568, 1-569, 1-570, 1-571, 1-572, 1-573, 1-574, 1-575, 1-576, 1-577, 1-578, 1-579, 1-580, 1-581, 1-582, 1-583, 1-584, 1-585, 1-586, 1-587, 1-588, 1-589, 1-590, 1-591, 1-592, 1-593, 1-594, 1-595, 1-596, 1-597, 1-598, 1-599, 1-600, 1-601, 1-602, 1-603, 1-604, 1-605, 1-606, 1-607, 1-608, 1-609, 1-610, 1-611, 1-612, 1-613, 1-614, 1-615, 1-616, 1-617, 1-618, 1-619, 1-620, 1-621, 1-622, 1-623, 1-624, 1-625, 1-626, 1-627, 1-628, 1-629, 1-630, 1-631, 1-632, 1-633, 1-634, 1-635, 1-636, 1-637, 1-638, 1-639, 1-640, 1-641, 1-642, 1-643, 1-644, 1-645, 1-646, 1-647, 1-648, 1-649, or 1-650 in SEQ ID NO: 1. In some embodiments, the N-terminal portion of the split Cas9 protein comprises a portion of any one of SEQ ID NOs: 1-275 and 394-397 that corresponds to amino acids 1-573 or 1-637 of SEQ ID NO: 1.

[0099] The C-terminal portion of the split Cas9 can be joined with the N-terminal portion of the split Cas9 to form a complete Cas9 protein. In some embodiments, the C-terminal portion of the Cas9 protein starts from where the N-terminal portion of the Cas9 protein ends. As such, in some embodiments, the C-terminal portion of the split Cas9 comprises a portion of any one of SEQ ID NO: 1-275 and 394-397 that corresponds to amino acids (551-651)-1368 of SEQ ID NO: 1. "(551-651)-1368" means starting at an amino acid between amino acids 551-651 (inclusive) and ending at amino acid 1368. For example, the C-terminal portion of the split Cas9 may comprise a

portion of any one of SEQ ID NO: 1-275 and 394-397 that corresponds to amino acid 551-1368, 552-1368, 553-1368, 554-1368, 555-1368, 556-1368, 557-1368, 558-1368, 559-1368, 560-1368, 561-1368, 562-1368, 563-1368, 564-1368, 565-1368, 566-1368, 567-1368, 568-1368, 569-1368, 570-1368, 571-1368, 572-1368, 573-1368, 574-1368, 575-1368, 576-1368, 577-1368, 578-1368, 579-1368, 580-1368, 581-1368, 582-1368, 583-1368, 584-1368, 585-1368, 586-1368, 587-1368, 588-1368, 589-1368, 590-1368, 591-1368, 592-1368, 593-1368, 594-1368, 595-1368, 596-1368, 597-1368, 598-1368, 599-1368, 600-1368, 601-1368, 602-1368, 603-1368, 604-1368, 605-1368, 606-1368, 607-1368, 608-1368, 609-1368, 610-1368, 611-1368, 612-1368, 613-1368, 614-1368, 615-1368, 616-1368, 617-1368, 618-1368, 619-1368, 620-1368, 621-1368, 622-1368, 623-1368, 624-1368, 625-1368, 626-1368, 627-1368, 628-1368, 629-1368, 630-1368, 631-1368, 632-1368, 633-1368, 634-1368, 635-1368, 636-1368, 637-1368, 638-1368, 639-1368, 640-1368, 641-1368, 642-1368, 643-1368, 644-1368, 645-1368, 646-1368, 647-1368, 648-1368, 649-1368, 650-1368, or 651-1368 of SEQ ID NO: 1. In some embodiments, the C-terminal portion of the split Cas9 protein comprises a portion of any one of SEQ ID NO: 1-275 and 394-397 that corresponds to amino acids 574-1368 or 638-1368 of SEQ ID NO: 1.

[00100] Cas9 variants may also be delivered to cells using the methods described herein. For example, a Cas9 variant may also be "split" as described herein. A Cas9 variant may comprise an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of the Cas9 sequences provided herein. In some embodiments, the Cas9 variant comprises an amino acid sequence that is shorter or longer in length (*e.g.*, by no more than 30%, no more than 25%, no more than 20%, no more than 15%, no more than 10%, no more than 5%, no more than 1% longer or shorter) than any of the Cas9 proteins provided herein (*e.g.*, in Example 1). In some embodiments, the UGI comprises an amino acid sequence that is shorter or longer in length (*e.g.*, by no more than 200 amino acids, no more than 150 amino acids, no more than 100 amino acids, no more than 50 amino acids, no more than 10 amino acids, no more than 5 amino acids, or no more than 2 amino acids longer or shorter) than any of the Cas9 proteins provided herein.

[00101] In some embodiments, the N-terminal portion of a split Cas9 comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least

99.5% identical to the corresponding portion of any one of the Cas9 sequences provided herein (*e.g.*, in Example 1). In some embodiments, the N-terminal portion of the split Cas9 comprises an amino acid sequence that is shorter or longer in length (*e.g.*, by no more than 30%, no more than 25%, no more than 20%, no more than 15%, no more than 10%, no more than 5%, no more than 1% longer or shorter) than the corresponding portion of any of the Cas9 proteins provided herein. In some embodiments, the N-terminal portion of the split Cas9 comprises an amino acid sequence that is shorter or longer in length (*e.g.*, by no more than 200 amino acids, no more than 150 amino acids, no more than 100 amino acids, no more than 50 amino acids, no more than 10 amino acids, no more than 5 amino acids, or no more than 2 amino acids longer or shorter) than the corresponding portion of any of the Cas9 proteins provided herein.

[00102] In some embodiments, the C-terminal portion of a split Cas9 comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to the corresponding portion of any one of the Cas9 sequences provided herein (*e.g.*, in Example 1). In some embodiments, the C-terminal portion of the split Cas9 comprises an amino acid sequence that is shorter or longer in length (*e.g.*, by no more than 30%, no more than 25%, no more than 20%, no more than 15%, no more than 10%, no more than 5%, no more than 1% longer or shorter) than the corresponding portion of any of the Cas9 proteins provided herein. In some embodiments, the C-terminal portion of the split Cas9 comprises an amino acid sequence that is shorter or longer in length (*e.g.*, by no more than 200 amino acids, no more than 150 amino acids, no more than 100 amino acids, no more than 50 amino acids, no more than 10 amino acids, no more than 5 amino acids, or no more than 2 amino acids longer or shorter) than the corresponding portion of any of the Cas9 proteins provided herein.

[00103] In some embodiments, the Cas9 variant is a dCas9 or nCas9. In some embodiments, the N-terminal portion of the split Cas9 comprises a mutation corresponding to a D10A mutation in SEQ ID NO: 1. In some embodiments, the N-terminal portion of the split Cas9 comprises a mutation corresponding to a D10A mutation in SEQ ID NO: 1 and the C-terminal portion of the split Cas9 comprises a mutation corresponding to a H840A mutation in SEQ ID NO:1. In some embodiments, the N-terminal portion of the split Cas9 comprises a mutation corresponding to a D10A mutation in SEQ ID NO: 1, and the C-terminal portion of the split Cas9 comprises a histidine at the position corresponding to position 840 in SEQ ID NO:1.

[00104] In some embodiments, to join the N-terminal portion of the Cas9 protein and the C-terminal portion of the Cas9 protein, an intein system may be used. In some embodiments, the N-terminal portion of the Cas9 is fused to an intein-N. In some embodiments, the intein-N is fused to the C-terminus of the N-terminal portion of the Cas9 to form a structure of NH_2-[N-terminal portion of Cas9]-[intein-N]-COOH. In some embodiments, the intein-N is encoded by the dnaE-n gene. In some embodiments, the intein-N comprises the amino acid sequence of any one of SEQ ID NOs: 350-351 and 354-355. In some embodiments, the C-terminal portion of the Cas9 is fused to an intein-C, and the intein-C is fused to the N-terminus of the C-terminal portion of the Cas9 to form a structure of NH_2-[intein-C]-[C-terminal portion of Cas9]-COOH. In some embodiments, the intein-C is encoded by the dnaE-c gene. In some embodiments, the intein-C comprises the amino acid sequence of any one of SEQ ID NOs: 352-353 and 356-357. Other split intein systems may also be used in the present disclosure and are known in the art.

[00105] Split nucleobase editors may be used in the present disclosure. Some aspects of the present disclosure relate to compositions comprising (i) a first nucleotide sequence encoding an N-terminal portion of a nucleobase editor fused at its C-terminus to an intein-N; and (ii) a second nucleotide sequence encoding an intein-C fused to the N-terminus of a C-terminal portion of the nucleobase editor.

[00106] Nucleobase editor variants are contemplated. For example, a nucleobase editor variant may also be "split" as described herein. A nucleobase editor variant may comprise an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of the nucleobase editor sequences (SEQ ID NOs: X-X) provided herein.

[00107] In some embodiments, the N-terminal portion of a split nucleobase editor comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to the corresponding portion of any one of the nucleobase editors provided herein (*e.g.*, in Example 1). In some embodiments, the N-terminal portion of the split nucleobase editor comprises an amino acid sequence that is shorter or longer in length (*e.g.*, by no more than 30%, no more than 25%, no more than 20%, no more than 15%, no more than 10%, no more than 5%, no more than 1% longer or shorter) than the corresponding portion of any of the

nucleobase editors provided herein. In some embodiments, the N-terminal portion of the split nucleobase editor comprises an amino acid sequence that is shorter or longer in length (*e.g.*, by no more than 200 amino acids, no more than 150 amino acids, no more than 100 amino acids, no more than 50 amino acids, no more than 10 amino acids, no more than 5 amino acids, or no more than 2 amino acids longer or shorter) than the corresponding portion of any of the nucleobase editors provided herein.

[00108] In some embodiments, the C-terminal portion of a split nucleobase editor comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to the corresponding portion of any one of the nucleobase editors provided herein (*e.g.*, in Example 1). In some embodiments, the C-terminal portion of the split nucleobase editor comprises an amino acid sequence that is shorter or longer in length (*e.g.*, by no more than 30%, no more than 25%, no more than 20%, no more than 15%, no more than 10%, no more than 5%, no more than 1% longer or shorter) than the corresponding portion of any of the nucleobase editors provided herein. In some embodiments, the C-terminal portion of the split nucleobase editor comprises an amino acid sequence that is shorter or longer in length (*e.g.*, by no more than 200 amino acids, no more than 150 amino acids, no more than 100 amino acids, no more than 50 amino acids, no more than 10 amino acids, no more than 5 amino acids, or no more than 2 amino acids longer or shorter) than the corresponding portion of any of the nucleobase editors provided herein.

[00109] As described herein, the N-terminal portion of a nucleobase editor comprises the N-terminal portion of a nuclease-inactive Cas9 protein (dCas9) or a Cas9 nickase (nCas9) . In some embodiments, the N-terminal portion of a nucleobase editor further comprises a nucleobase modifying enzyme (*e.g.*, nucleases, nickases, recombinases, deaminases, DNA repair enzymes, DNA damage enzymes, dismutases, alkylation enzymes, depurination enzymes, oxidation enzymes, pyrimidine dimer forming enzymes, integrases, transposases, polymerases, ligases, helicases, photolyases, glycosylases, epigenetic modifiers such as methylases, acetylases, methyltransferase, demethylase, *etc.*). In some embodiments, the nucleobase modifying enzyme is a deaminase (*e.g.*, a cytosine deaminase or an adenosine deaminase, or functional variants thereof). In some embodiments, the nucleobase modifying enzyme is fused to the N-terminus of the N-terminal portion of the split dCas9 or split nCas9. In some embodiments, the N-terminal

portion of the nucleobase editor has of the structure: NH_2-[nucleobase modifying enzyme]-[N-terminal portion of dCas9 or nCas9]-COOH. In some embodiments, the N-terminal portion of the nucleobase editor is fused to an intein N. In some embodiments, the intein-N is fused to the C-terminus of the N-terminal portion of the nucleobase editor.

[00110] In some embodiments, the first nucleotide sequence encodes a polypeptide comprising the structure NH_2-[nucleobase modifying enzyme]-[N-terminal portion of dCas9 or nCas9]-[intein-N]-COOH.

[00111] In some embodiments, the C-terminal portion of the nucleobase editor comprises the C-terminal portion of a nuclease-inactive Cas9 protein (dCas9) or a Cas9 nickase (nCas9). In some embodiments, the nucleobase modifying enzyme is fused to the C-terminus of the C-terminal portion of the split dCas9 or split nCas9. In some embodiments, the C-terminal portion of the nucleobase editoris of the structure: NH_2-[C-terminal portion of dCas9 or nCas9]-[nucleobase modifying enzyme]-COOH. In some embodiments, the C-terminal portion of the nucleobase editor comprises an intein-C fused to the C-terminal portion of the Cas9 protein. In some embodiments, the intein-C is fused to the N-terminus of the C-terminal portion of the nucleobase editor. In some embodiments, the second nucleotide sequence encodes a polypeptide of the structure: NH_2-[intein-C]-[C-terminal portion of the Cas9 protein]-COOH.

[00112] In some embodiments, the N-terminal portion of a split nucleobase editor further comprises an inhibitor of uracil glycosylase (UGI). In some embodiments, the first nucleotide sequence encodes a polypeptide of the structure: NH_2-[UGI]-[nucleobase modifying enzyme]-[N-terminal portion of dCas9 or nCas9]-[intein-N]. In some embodiments, the first nucleotide sequence encodes a polypeptide is of the structure: NH_2-[nucleobase modifying enzyme]-[UGI]-[N-terminal portion of dCas9 or nCas9]-[intein-N].

[00113] In some embodiments, the C-terminal portion of a split nucleobase editor further comprises an enzyme that inhibits the activity of uracil glycosylase (UGI). In some embodiments, the second nucleotide sequence encodes a polypeptide of the structure: NH_2-[intein-C]-[C-terminal portion of dCas9 or nCas9]-[UGI]-COOH. In some embodiments, the second nucleotide sequence encodes a polypeptide of the structure: NH_2-[intein-C]-[C-terminal portion of dCas9 or nCas9]-[nucleobase modifying enzyme]-[UGI]-COOH. In some embodiments, the second nucleotide sequence encodes a polypeptide of the structure: NH_2-

[intein-C]-[C-terminal portion of dCas9 or nCas9]-[UGI]-[nucleobase modifying enzyme]-COOH.

[00114] In some embodiments, when the N-terminal portion and the C-terminal portion of the nucleobase are joined, to form a complete split nucleobase editor. In some embodiments, the split nucleobase editor may comprise any one of the following structures:

NH_2-[nucleobase modifying enzyme]-[dCas9 or nCas9]-COOH

NH_2-[UGI]-[nucleobase modifying enzyme]-[dCas9 or nCas9]-COOH

NH_2-[nucleobase modifying enzyme]-[UGI]-[dCas9 or nCas9]-COOH

NH_2-[nucleobase modifying enzyme]-[dCas9 or nCas9]-[UGI]-COOH

NH_2-[dCas9 or nCas9]-[nucleobase modifying enzyme]-COOH

NH_2-[UGI]-[dCas9 or nCas9]-[nucleobase modifying enzyme]-COOH

NH_2-[dCas9 or nCas9]-[UGI]-[nucleobase modifying enzyme]-COOH or

NH_2-[dCas9 or nCas9]-[nucleobase modifying enzyme]-[UGI]-COOH.

[00115] In some embodiments, the first nucleotide sequence or the second nucleotide sequence (encoding either the split Cas9 protein or the split nucleobase editor) is operably linked to a nucleotide sequence encoding at least one bipartite nuclear localization signal (NLS). For example, the first nucleotide sequence may be operably linked to a nucleotide sequence encoding one or more (*e.g.*, 2, 3, 4, 5, or more) bipartite NLS. In some embodiments, the second nucleotide sequence may be operably linked to a nucleotide sequence encoding one or more (*e.g.*, 2, 3, 4, 5, or more) bipartite NLSs. As such, the split Cas9 or split nucleobase editor formed by joining the N-terminal portion and the C-terminal portion may comprise one or more bipartite NLSs. For example, the split Cas9 or split nucleobase editor may comprise any one of the following structures (bNLS means one or more bipartite nuclear localization signals):

NH_2-bNLS-[Cas9]-COOH

NH_2-[Cas9]-bNLS-COOH

NH_2-bNLS-[nucleobase modifying enzyme]-[dCas9 or nCas9]-COOH

NH_2-[nucleobase modifying enzyme]-bNLS-[dCas9 or nCas9]-COOH

NH_2-[nucleobase modifying enzyme]-[dCas9 or nCas9]-bNLS-COOH

NH_2-bNLS-[nucleobase modifying enzyme]-[dCas9 or nCas9]-bNLS-COOH

NH_2-bNLS-[nucleobase modifying enzyme]-[dCas9 or nCas9]-COOH

NH_2-bNLS-[nucleobase modifying enzyme]-bNLS-[dCas9 or nCas9]-bNLS-COOH

NH_2-bNLS-[UGI]-[nucleobase modifying enzyme]-[dCas9 or nCas9]-COOH

NH_2-[UGI]-bNLS-[nucleobase modifying enzyme]-[dCas9 or nCas9]-COOH

NH_2-[UGI]-[nucleobase modifying enzyme]-bNLS[dCas9 or nCas9]-COOH

NH_2-[UGI]-[nucleobase modifying enzyme]-[dCas9 or nCas9]-bNLS-COOH

NH_2-bNLS-[UGI]-bNLS-[nucleobase modifying enzyme]-[dCas9 or nCas9]-COOH

NH_2-bNLS-[UGI]-[nucleobase modifying enzyme]-bNLS-[dCas9 or nCas9]-COOH

NH_2-bNLS-[UGI]-[nucleobase modifying enzyme]-[dCas9 or nCas9]-bNLS-COOH

NH_2-[UGI]-bNLS-[nucleobase modifying enzyme]-bNLS-[dCas9 or nCas9]-COOH

NH_2-[UGI]-bNLS-[nucleobase modifying enzyme]-[dCas9 or nCas9]-bNLS-COOH

NH_2-[UGI]-[nucleobase modifying enzyme]-bNLS-[dCas9 or nCas9]-bNLS-COOH

NH_2-bNLS-[UGI]-bNLS-[nucleobase modifying enzyme]-bNLS-[dCas9 or nCas9]-COOH

NH_2-bNLS-[UGI]-[nucleobase modifying enzyme]-bNLS-[dCas9 or nCas9]-bNLS-COOH

NH_2-[UGI]-bNLS-[nucleobase modifying enzyme]-bNLS-[dCas9 or nCas9]-bNLS-COOH

NH_2-bNLS-[UGI]-bNLS-[nucleobase modifying enzyme]-bNLS-[dCas9 or nCas9]-bNLS-COOH

NH_2-bNLS-[nucleobase modifying enzyme]-[UGI]-[dCas9 or nCas9]-COOH

NH_2-[nucleobase modifying enzyme]-bNLS-[UGI]-[dCas9 or nCas9]-COOH

NH_2-[nucleobase modifying enzyme]-[UGI]-bNLS-[dCas9 or nCas9]-bNLS-COOH

NH_2-bNLS-[nucleobase modifying enzyme]-bNLS-[UGI]-[dCas9 or nCas9]-COOH

NH_2-bNLS-[nucleobase modifying enzyme]-[UGI]-bNLS-[dCas9 or nCas9]-COOH

NH_2-bNLS-[nucleobase modifying enzyme]-[UGI]-[dCas9 or nCas9]-bNLS-COOH

NH_2-[nucleobase modifying enzyme]-bNLS-[UGI]-bNLS-[dCas9 or nCas9]-COOH

NH_2-[nucleobase modifying enzyme]-bNLS-[UGI]-[dCas9 or nCas9]-bNLS-COOH

NH_2-[nucleobase modifying enzyme]-[UGI]-bNLS-[dCas9 or nCas9]-bNLS-COOH

NH_2-bNLS-[nucleobase modifying enzyme]-bNLS-[UGI]-bNLS-[dCas9 or nCas9]-COOH

NH_2-bNLS-[nucleobase modifying enzyme]-bNLS-[UGI]-[dCas9 or nCas9]-bNLS-COOH

NH_2-[nucleobase modifying enzyme]-bNLS-[UGI]-bNLS-[dCas9 or nCas9]-bNLS-COOH

NH_2-bNLS-[nucleobase modifying enzyme]-bNLS-[UGI]-bNLS-[dCas9 or nCas9]-bNLS-COOH

NH_2-bNLS-[nucleobase modifying enzyme]-[dCas9 or nCas9]-[UGI]-COOH

NH_2-[nucleobase modifying enzyme]-bNLS-[dCas9 or nCas9]-[UGI]-COOH

NH_2-[nucleobase modifying enzyme]-[dCas9 or nCas9]-bNLS-[UGI]-COOH

NH_2-[nucleobase modifying enzyme]-[dCas9 or nCas9]-[UGI]-bNLS-COOH

NH_2-bNLS-[nucleobase modifying enzyme]-bNLS-[dCas9 or nCas9]-[UGI]-COOH

NH_2-bNLS-[nucleobase modifying enzyme]-[dCas9 or nCas9]-bNLS-[UGI]-COOH

NH_2-bNLS-[nucleobase modifying enzyme]-[dCas9 or nCas9]-[UGI]-bNLS-COOH

NH_2-bNLS-[nucleobase modifying enzyme]-bNLS-[dCas9 or nCas9]-bNLS-[UGI]-COOH

NH_2-bNLS-[nucleobase modifying enzyme]-bNLS-[dCas9 or nCas9]-[UGI]-bNLS-COOH

NH_2-bNLS-[nucleobase modifying enzyme]-[dCas9 or nCas9]-bNLS-[UGI]-bNLS-COOH

NH_2-[nucleobase modifying enzyme]-bNLS-[dCas9 or nCas9]-bNLS-[UGI]-bNLS-COOH

NH_2-bNLS-[nucleobase modifying enzyme]-bNLS-[dCas9 or nCas9]-bNLS-[UGI]-bNLS-COOH

NH_2-bNLS-[dCas9 or nCas9]-[nucleobase modifying enzyme]-COOH

NH_2-[dCas9 or nCas9]-bNLS-[nucleobase modifying enzyme]-COOH

NH_2-[dCas9 or nCas9]-[nucleobase modifying enzyme]-bNLS-COOH

NH_2-bNLS-[dCas9 or nCas9]-bNLS-[nucleobase modifying enzyme]-COOH

NH_2-bNLS-[dCas9 or nCas9]-[nucleobase modifying enzyme]-bNLS-COOH

NH_2-[dCas9 or nCas9]-bNLS-[nucleobase modifying enzyme]-bNLS-COOH

NH_2-bNLS-[dCas9 or nCas9]-bNLS-[nucleobase modifying enzyme]-bNLS-COOH

NH_2-bNLS-[UGI]-[dCas9 or nCas9]-[nucleobase modifying enzyme]-COOH

NH_2-[UGI]-bNLS-[dCas9 or nCas9]-[nucleobase modifying enzyme]-COOH

NH_2-[UGI]-[dCas9 or nCas9]-bNLS-[nucleobase modifying enzyme]-COOH

NH_2-[UGI]-[dCas9 or nCas9]-[nucleobase modifying enzyme]-bNLS-COOH

NH_2-bNLS-[UGI]-bNLS-[dCas9 or nCas9]-[nucleobase modifying enzyme]-COOH

NH_2-bNLS-[UGI]-[dCas9 or nCas9]-bNLS-[nucleobase modifying enzyme]-COOH

NH_2-bNLS-[UGI]-[dCas9 or nCas9]-[nucleobase modifying enzyme]-bNLS-COOH

NH_2-[UGI]-bNLS-[dCas9 or nCas9]-bNLS-[nucleobase modifying enzyme]-COOH

NH_2-[UGI]-bNLS-[dCas9 or nCas9]-[nucleobase modifying enzyme]-bNLS-COOH

NH_2-[UGI]-[dCas9 or nCas9]-bNLS-[nucleobase modifying enzyme]-bNLS-COOH

NH_2-bNLS-[UGI]-bNLS-[dCas9 or nCas9]-bNLS-[nucleobase modifying enzyme]-COOH

NH_2-bNLS-[UGI]-[dCas9 or nCas9]-bNLS-[nucleobase modifying enzyme]-bNLS-COOH

NH_2-bNLS-[UGI]-bNLS-[dCas9 or nCas9]-[nucleobase modifying enzyme]-bNLS-COOH

NH_2-[UGI]-bNLS-[dCas9 or nCas9]-bNLS-[nucleobase modifying enzyme]-bNLS-COOH

NH_2-bNLS-[UGI]-bNLS-[dCas9 or nCas9]-bNLS-[nucleobase modifying enzyme]-bNLS-COOH

NH_2-bNLS-[dCas9 or nCas9]-[UGI]-[nucleobase modifying enzyme]-COOH

NH_2-[dCas9 or nCas9]-bNLS-[UGI]-[nucleobase modifying enzyme]-COOH

NH_2-[dCas9 or nCas9]-[UGI]-bNLS-[nucleobase modifying enzyme]-COOH

NH_2-[dCas9 or nCas9]-[UGI]-[nucleobase modifying enzyme]-bNLS-COOH

NH_2-bNLS-[dCas9 or nCas9]-bNLS[UGI]-[nucleobase modifying enzyme]-COOH

NH_2-bNLS-[dCas9 or nCas9]-[UGI]-bNLS-[nucleobase modifying enzyme]-COOH

NH_2-bNLS-[dCas9 or nCas9]-[UGI]-[nucleobase modifying enzyme]-bNLS-COOH

NH_2-[dCas9 or nCas9]-bNLS-[UGI]-bNLS-[nucleobase modifying enzyme]-COOH

NH_2-[dCas9 or nCas9]-bNLS-[UGI]-[nucleobase modifying enzyme]-bNLS-COOH

NH_2-[dCas9 or nCas9]-[UGI]-bNLS-[nucleobase modifying enzyme]-bNLS-COOH

NH_2-bNLS-[dCas9 or nCas9]-bNLS-[UGI]-bNLS-[nucleobase modifying enzyme]-COOH

NH_2-bNLS-[dCas9 or nCas9]-bNLS-[UGI]-[nucleobase modifying enzyme]-bNLS-COOH

NH_2-bNLS-[dCas9 or nCas9]-[UGI]-bNLS-[nucleobase modifying enzyme]-bNLS-COOH

NH_2-[dCas9 or nCas9]-bNLS-[UGI]-bNLS-[nucleobase modifying enzyme]-bNLS-COOH

NH_2-bNLS-[dCas9 or nCas9]-bNLS-[UGI]-bNLS-[nucleobase modifying enzyme]-bNLS-COOH

NH_2-bNLS-[dCas9 or nCas9]-[nucleobase modifying enzyme]-[UGI]-COOH

NH_2-[dCas9 or nCas9]-bNLS-[nucleobase modifying enzyme]-[UGI]-COOH

NH_2-[dCas9 or nCas9]-[nucleobase modifying enzyme]-bNLS-[UGI]-COOH

NH_2-[dCas9 or nCas9]-[nucleobase modifying enzyme]-[UGI]-bNLS-COOH

NH_2-bNLS-[dCas9 or nCas9]-bNLS-[nucleobase modifying enzyme]-[UGI]-COOH

NH_2-bNLS-[dCas9 or nCas9]-[nucleobase modifying enzyme]-bNLS-[UGI]-COOH

NH_2-bNLS-[dCas9 or nCas9]-[nucleobase modifying enzyme]-[UGI]-bNLS-COOH

NH_2-[dCas9 or nCas9]-bNLS-[nucleobase modifying enzyme]-bNLS-[UGI]-COOH

NH_2-[dCas9 or nCas9]-bNLS-[nucleobase modifying enzyme]-[UGI]-bNLS-COOH

NH_2-[dCas9 or nCas9]-[nucleobase modifying enzyme]-bNLS-[UGI]-bNLS-COOH

NH_2-bNLS-[dCas9 or nCas9]-bNLS-[nucleobase modifying enzyme]-bNLS-[UGI]-COOH

NH_2-bNLS-[dCas9 or nCas9]-bNLS-[nucleobase modifying enzyme]-[UGI]-bNLS-COOH

NH_2-bNLS-[dCas9 or nCas9]-[nucleobase modifying enzyme]-bNLS-[UGI]-bNLS-COOH

NH_2-[dCas9 or nCas9]-bNLS-[nucleobase modifying enzyme]-bNLS-[UGI]-bNLS-COOH or

NH_2-bNLS-[dCas9 or nCas9]-bNLS-[nucleobase modifying enzyme]-bNLS-[UGI]-bNLS-COOH

[00116] Herein, "NH_2-" represents the N-terminus of a protein or polypeptide, and "-COOH" represents the C-terminus of a protein or polypeptide. "]-[" represents a peptide bond or a linker. In some embodiments, linkers may be used to link any of the protein or protein domains described herein. The linker may be as simple as a covalent bond, or it may be a polymeric linker many atoms in length. In some embodiments, the linker is a polypeptide or based on

amino acids. In some embodiments, the linker is not peptide-like. In some embodiments, the linker is a covalent bond (*e.g.*, a carbon-carbon bond, disulfide bond, carbon-heteroatom bond, etc.). In some embodiments, the linker is a carbon-nitrogen bond of an amide linkage. In some embodiments, the linker is a cyclic or acyclic, substituted or unsubstituted, branched or unbranched aliphatic or heteroaliphatic linker. In some embodiments, the linker is polymeric (*e.g.*, polyethylene, polyethylene glycol, polyamide, polyester, etc.). In some embodiments, the linker comprises a monomer, dimer, or polymer of aminoalkanoic acid. In some embodiments, the linker comprises an aminoalkanoic acid (*e.g.*, glycine, ethanoic acid, alanine, beta-alanine, 3-aminopropanoic acid, 4-aminobutanoic acid, 5-pentanoic acid, etc.). In some embodiments, the linker comprises a monomer, dimer, or polymer of aminohexanoic acid (Ahx). In some embodiments, the linker is based on a carbocyclic moiety (*e.g.*, cyclopentane, cyclohexane). In some embodiments, the linker comprises a polyethylene glycol moiety (PEG). In some embodiments, the linker comprises amino acids. In some embodiments, the linker comprises a peptide. In some embodiments, the linker comprises an aryl or heteroaryl moiety. In some embodiments, the linker is based on a phenyl ring. The linker may include functionalized moieties to facilitate attachment of a nucleophile (*e.g.*, thiol, amino) from the peptide to the linker. Any electrophile may be used as part of the linker. Exemplary electrophiles include, but are not limited to, activated esters, activated amides, Michael acceptors, alkyl halides, aryl halides, acyl halides, and isothiocyanates.

[00117] In some embodiments, the linker is an amino acid or a plurality of amino acids (*e.g.*, a peptide or protein). In some embodiments, the linker is a bond (*e.g.*, a covalent bond), an organic molecule, group, polymer, or chemical moiety. In some embodiments, the linker is 5-100 amino acids in length, for example, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-110, 110-120, 120-130, 130-140, 140-150, or 150-200 amino acids in length. Longer or shorter linkers are also contemplated. In some embodiments, a linker comprises the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 377), which may also be referred to as the XTEN linker. In some embodiments, a linker comprises the amino acid sequence: SGGS (SEQ ID NO: 378). In some embodiments, a linker comprises the amino acid sequence: $(SGGS)_n$ (SEQ ID NO: 379), $(GGGS)_n$ (SEQ ID NO: 380), $(GGGGS)_n$ (SEQ ID NO: 381), $(G)_n$ (SEQ ID NO: 390), $(EAAAK)_n$ (SEQ ID NO: 382), $(GGS)_n$, SGSETPGTSESATPES (SEQ ID NO: 377),

or $(XP)_n$ motif, or a combination of any of these, wherein n is independently an integer between 1 and 30, inclusive, and wherein X is any amino acid. In some embodiments, n is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15. In some embodiments, the linker comprises the amino acid sequence: SGSETPGTSESATPES (SEQ ID NO: 377), and SGGS (SEQ ID NO: 378). In some embodiments, the linker comprises the amino acid sequence: SGGSSGSETPGTSESATPESSGGS (SEQ ID NO: 383). In some embodiments, a linker comprises the amino acid sequence: SGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID NO: 384). In some embodiments, a linker comprises the amino acid sequence: GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS (SEQ ID NO: 385).

[00118] In some embodiments, the linker is 24 amino acids in length. In some embodiments, the linker comprises the amino acid sequence SGGSSGGSSGSETPGTSESATPES (SEQ ID NO: 343). In some embodiments, the linker is 40 amino acids in length. In some embodiments, the linker comprises the amino acid sequence SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGS (SEQ ID NO: 391). In some embodiments, the linker is 64 amino acids in length. In some embodiments, the linker comprises the amino acid sequence SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID NO: 392). In some embodiments, the linker is 92 amino acids in length. In some embodiments, the linker comprises the amino acid sequence PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATS (SEQ ID NO: 393).

[00119] In some embodiments, the first and second nucleotide sequences are on the same nucleic acid vector. In some embodiments, the first and second nucleotide sequences are on different nucleic acid vectors. In some embodiments, the vector is a plasmid. In some embodiments, the nucleic acid vector is a recombinant genome of a adeno-associated virus (rAAV). In some embodiments, the nucleic acid vector is the genome of an adeno-associated virus packaged in a rAAV particle. In some embodiments, the first and/or the second nucleotide sequence is operably linked to a promoter. In some embodiments, the nucleic acid vector further comprise a nucleotide sequence encoding one or more (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more) gRNAs

operably linked to a promoter. In some embodiments, the promoter is a constitutive promoter. In some embodiments, the promoter is an inducible promoter.

[00120] An inducible promoter of the present disclosure may be induced by (or repressed by) one or more physiological condition(s), such as changes in light, pH, temperature, radiation, osmotic pressure, saline gradients, cell surface binding, and the concentration of one or more extrinsic or intrinsic inducing agent(s). An extrinsic inducer signal or inducing agent may comprise, without limitation, amino acids and amino acid analogs, saccharides and polysaccharides, nucleic acids, protein transcriptional activators and repressors, cytokines, toxins, petroleum-based compounds, metal containing compounds, salts, ions, enzyme substrate analogs, hormones, or combinations thereof.

[00121] Inducible promoters of the present disclosure include any inducible promoter described herein or known to one of ordinary skill in the art. Examples of inducible promoters include, without limitation, chemically/biochemically-regulated and physically-regulated promoters such as alcohol-regulated promoters, tetracycline-regulated promoters (*e.g.*, anhydrotetracycline (aTc)-responsive promoters and other tetracycline-responsive promoter systems, which include a tetracycline repressor protein (tetR), a tetracycline operator sequence (tetO) and a tetracycline transactivator fusion protein (tTA)), steroid-regulated promoters (*e.g.*, promoters based on the rat glucocorticoid receptor, human estrogen receptor, moth ecdysone receptors, and promoters from the steroid/retinoid/thyroid receptor superfamily), metal-regulated promoters (*e.g.*, promoters derived from metallothionein (proteins that bind and sequester metal ions) genes from yeast, mouse and human), pathogenesis-regulated promoters (*e.g.*, induced by salicylic acid, ethylene or benzothiadiazole (BTH)), temperature/heat-inducible promoters (*e.g.*, heat shock promoters), and light-regulated promoters (*e.g.*, light responsive promoters from plant cells). Other inducible promoter systems are known in the art and may be used in accordance with the present disclosure.

[00122] In some embodiments, inducible promoters of the present disclosure function in prokaryotic cells (*e.g.*, bacterial cells). Examples of inducible promoters for use prokaryotic cells include, without limitation, bacteriophage promoters (*e.g.* Pls1con, T3, T7, SP6, PL) and bacterial promoters (*e.g.*, Pbad, PmgrB, Ptrc2, Plac/ara, Ptac, Pm), or hybrids thereof (*e.g.* PLlacO, PLtetO). Examples of bacterial promoters for use in accordance with the present disclosure include, without limitation, positively regulated *E. coli* promoters, such as positively

regulated σ70 promoters (*e.g.*, inducible pBad/araC promoter, Lux cassette right promoter, modified lamdba Prm promote, plac Or2-62 (positive), pBad/AraC with extra REN sites, pBad, P(Las) TetO, P(Las) CIO, P(Rhl), Pu, FecA, pRE, cadC, hns, pLas, pLux), σS promoters (*e.g.*, Pdps), σ32 promoters (*e.g.*, heat shock), and σ54 promoters (*e.g.*, glnAp2); negatively regulated *E. coli* promoters such as negatively regulated σ70 promoters (*e.g.*, Promoter (PRM+), modified lamdba Prm promoter, TetR - TetR-4C P(Las) TetO, P(Las) CIO, P(Lac) IQ, RecA_DlexO_DLacO1, dapAp, FecA, Pspac-hy, pcI, plux-cI, plux-lac, CinR, CinL, glucose controlled, modified Pr, modified Prm+, FecA, Pcya, rec A (SOS), Rec A (SOS), EmrR_regulated, BetI_regulated, pLac_lux, pTet_Lac, pLac/Mnt, pTet/Mnt, LsrA/cI, pLux/cI, LacI, LacIQ, pLacIQ1, pLas/cI, pLas/Lux, pLux/Las, pRecA with LexA binding site, reverse BBa_R0011, pLacI/ara-1, pLacIq, rrnB P1, cadC, hns, PfhuA, pBad/araC, nhaA, OmpF, RcnR), σS promoters (*e.g.*, Lutz-Bujard LacO with alternative sigma factor σ38), σ32 promoters (*e.g.*, Lutz-Bujard LacO with alternative sigma factor σ32), and σ54 promoters (*e.g.*, glnAp2); negatively regulated B. subtilis promoters such as repressible B. subtilis σA promoters (*e.g.*, Gram-positive IPTG-inducible, Xyl, hyper-spank) and σB promoters. Other inducible microbial promoters may be used in accordance with the present disclosure.

[00123] In some embodiments, inducible promoters of the present disclosure function in eukaryotic cells (*e.g.*, mammalian cells). Examples of inducible promoters for use eukaryotic cells include, without limitation, chemically-regulated promoters (*e.g.*, alcohol-regulated promoters, tetracycline-regulated promoters, steroid-regulated promoters, metal-regulated promoters, and pathogenesis-related (PR) promoters) and physically-regulated promoters (*e.g.*, temperature-regulated promoters and light-regulated promoters).

Recombinant Adeno-associated Virus (rAAV)

[00124] Some aspects of the present disclosure relate to using recombinant adeno-associated virus vectors for the delivery of a split Cas9 protein or a split nucleobase editor into a cell. The N-terminal portion of the Cas9 protein or the nucleobase editor and the C-terminal portion of the Cas9 protein or the nucleobase editor are delivered by separate rAAV vectors or particles into the same cell, since the full-length Cas9 protein or nucleobase editors exceeds the packaging limit of rAAV (~4.9 kb).

[00125] As such, in some embodiments, a composition for delivering the split Cas9 protein or split nucleobase editor into a cell (*e.g.*, a mammalian cell, a human cell) is provided. In some embodiments, the composition of the present disclosure comprises: (i) a first recombinant adeno-associated virus (rAAV) particle comprising a first nucleotide sequence encoding a N-terminal portion of a Cas9 protein or nucleobase editor fused at its C-terminus to an intein-N; and (ii) a second recombinant adeno-associated virus (rAAV) particle comprising a second nucleotide sequence encoding an intein-C fused to the N-terminus of a C-terminal portion of the Cas9 protein or nucleobase editor. The rAAV particles of the present disclosure comprise a rAAV vector (*i.e.*, a recombinant genome of the rAAV) encapsidated in the viral capsid proteins.

[00126] In some embodiments, the rAAV vector comprises: (1) a heterologous nucleic acid region comprising the first or second nucleotide sequence encoding the N-terminal portion or C-terminal portion of a split Cas9 protein or a split nucleobase editor in any form as described herein, (2) one or more nucleotide sequences comprising a sequence that facilitates expression of the heterologous nucleic acid region (*e.g.*, a promoter), and (3) one or more nucleic acid regions comprising a sequence that facilitate integration of the heterologous nucleic acid region (optionally with the one or more nucleic acid regions comprising a sequence that facilitates expression) into the genome of a cell. In some embodiments, viral sequences that facilitate integration comprise Inverted Terminal Repeat (ITR) sequences. In some embodiments, the first or second nucleotide sequence encoding the N-terminal portion or C-terminal portion of a split Cas9 protein or a split nucleobase editor is flanked on each side by an ITR sequence. In some embodiments, the nucleic acid vector further comprises a region encoding an AAV Rep protein as described herein, either contained within the region flanked by ITRs or outside the region. The ITR sequences can be derived from any AAV serotype (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) or can be derived from more than one serotype. In some embodiments, the ITR sequences are derived from AAV2 or AAV6.

[00127] ITR sequences and plasmids containing ITR sequences are known in the art and commercially available (see, *e.g.*, products and services available from Vector Biolabs, Philadelphia, PA; Cellbiolabs, San Diego, CA; Agilent Technologies, Santa Clara, Ca; and Addgene, Cambridge, MA; and Gene delivery to skeletal muscle results in sustained expression and systemic delivery of a therapeutic protein. Kessler PD, Podsakoff GM, Chen X, McQuiston SA, Colosi PC, Matelis LA, Kurtzman GJ, Byrne BJ. Proc Natl Acad Sci USA. 1996 Nov

26;93(24):14082-7; and Curtis A. Machida. Methods in Molecular Medicine™. Viral Vectors for Gene Therapy Methods and Protocols. 10.1385/1-59259-304-6:201 © Humana Press Inc. 2003. Chapter 10. Targeted Integration by Adeno-Associated Virus. Matthew D. Weitzman, Samuel M. Young Jr., Toni Cathomen and Richard Jude Samulski; U.S. Pat. Nos. 5,139,941 and 5,962,313, all of which are incorporated herein by reference). Exemplary ITR sequences are provided below.

AAV2:
TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTC GCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGA GGGAGTGGCCAACTCCATCACTAGGGGTTCCT (SEQ ID NO: 386)

AAV3:
TTGGCCACTCCCTCTATGCGCACTCGCTCGCTCGGTGGGGCCTGGCGACCAAAGGTC GCCAGACGGACGTGCTTTGCACGTCCGGCCCCACCGAGCGAGCGAGTGCGCATAGA GGGAGTGGCCAACTCCATCACTAGAGGTATGGC (SEQ ID NO: 387)

AAV5:
CTCTCCCCCCTGTCGCGTTCGCTCGCTCGCTGGCTCGTTTGGGGGGGTGGCAGCTCA AAGAGCTGCCAGACGACGGCCCTCTGGCCGTCGCCCCCCCAAACGAGCCAGCGAGC GAGCGAACGCGACAGGGGGGAGAGTGCCACACTCTCAAGCAAGGGGGTTTTGTA (SEQ ID NO: 388)

AAV6:
TTGCCCACTCCCTCTATGCGCGCTCGCTCGCTCGGTGGGGCCTGCGGACCAAAGGTC CGCAGACGGCAGAGCTCTGCTCTGCCGGCCCCACCGAGCGAGCGAGCGCGCATAGA GGGAGTGGGCAACTCCATCACTAGGGGTA (SEQ ID NO: 389)

[00128] In some embodiments, the rAAV vector of the present disclosure comprises one or more regulatory elements to control the expression of the heterologous nucleic acid region (*e.g.*, promoters, transcriptional terminators, and/or other regulatory elements). In some embodiments, the first and/or second nucleotide sequence is operably linked to one or more (*e.g.*, 1, 2, 3, 4, 5,

or more) transcriptional terminators. Non-limiting examples of transcriptional terminators that may be used in accordance with the present disclosure include transcription terminators of the bovine growth hormone gene (bGH), human growth hormone gene (hGH), SV40, CW3, ϕ, or combinations thereof. The efficiencies of several transcriptional terminators have been tested to determine their respective effects in the expression level of the split Cas9 protein or the split nucleobase editor (*e.g.*, see Figure 4). In some embodiments, the transcriptional terminator used in the present disclosure is a bGH transcriptional terminator. In some embodiments, the rAAV vector further comprises a Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element (WPRE). In some embodiments, the WPRE is inserted 5′ of the transcriptional terminator.

[00129] In some embodiments, the composition comprising the rAAV particle (in any form contemplated herein) further comprises a pharmaceutically acceptable carrier. In some embodiments, the composition is formulated in appropriate pharmaceutical vehicles for administration to human or animal subjects.

[00130] Some examples of materials which can serve as pharmaceutically-acceptable carriers include: (1) sugars, such as lactose, glucose and sucrose; (2) starches, such as corn starch and potato starch; (3) cellulose, and its derivatives, such as sodium carboxymethyl cellulose, methylcellulose, ethyl cellulose, microcrystalline cellulose and cellulose acetate; (4) powdered tragacanth; (5) malt; (6) gelatin; (7) lubricating agents, such as magnesium stearate, sodium lauryl sulfate and talc; (8) excipients, such as cocoa butter and suppository waxes; (9) oils, such as peanut oil, cottonseed oil, safflower oil, sesame oil, olive oil, corn oil and soybean oil; (10) glycols, such as propylene glycol; (11) polyols, such as glycerin, sorbitol, mannitol and polyethylene glycol (PEG); (12) esters, such as ethyl oleate and ethyl laurate; (13) agar; (14) buffering agents, such as magnesium hydroxide and aluminum hydroxide; (15) alginic acid; (16) pyrogen-free water; (17) isotonic saline; (18) Ringer's solution; (19) ethyl alcohol; (20) pH buffered solutions; (21) polyesters, polycarbonates and/or polyanhydrides; (22) bulking agents, such as polypeptides and amino acids (23) serum component, such as serum albumin, HDL and LDL; (22) C2-C12 alcohols, such as ethanol; and (23) other non-toxic compatible substances employed in pharmaceutical formulations. Wetting agents, coloring agents, release agents, coating agents, sweetening agents, flavoring agents, perfuming agents, preservative and antioxidants can also be present in the formulation. The terms such as "excipient", "carrier", "pharmaceutically acceptable carrier" or the like are used interchangeably herein.

Methods of Use

[00131] Other aspects of the present disclosure provide methods of delivering the split Cas9 protein or the split nucleobase editor into a cell to form a complete and functional Cas9 protein or nucleobase editor. For example, in some embodiments, a cell is contacted with a composition described herein (*e.g.*, compositions comprising nucleotide sequences encoding the split Cas9 or the split nucleobase editor or AAV particles containing nucleic acid vectors comprising such nucleotide sequences). In some embodiments, the contacting results in the delivery of such nucleotide sequences into a cell, wherein the N-terminal portion of the Cas9 protein or the nucleobase editor and the C-terminal portion of the Cas9 protein or the nucleobase editor are expressed in the cell and are joined to form a complete Cas9 protein or a complete nucleobase editor.

[00132] The split Cas9 protein or split nucleobase editor delivered using the methods described herein preferably have comparable activity compared to the original Cas9 protein or nucleobase editor (*i.e.*, unsplit protein delivered to a cell or expressed in a cell as a whole). For example, the split Cas9 protein or split nucleobase editor retains at least 50% (*e.g.*, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) of the activity of the original Cas9 protein or nucleobase editor. In some embodiments, the split Cas9 protein or split nucleobase editor is more active (*e.g.*, 2-fold, 5-fold, 10-fold, 100-fold, 1000-fold, or more) than that of an original Cas9 protein or nucleobase editor.

[00133] The compositions described herein may be administered to a subject in need thereof in a therapeutically effective amount to treat and/or prevent a disease or disorder the subject is suffering from. Any disease or disorder that maybe treated and/or prevented using CRISPR/Cas9-based genome-editing technology may be treated by the split Cas9 protein or the split nucleobase editor described herein. It is to be understood that, if the nucleotide sequences encoding the split Cas9 protein or the nucleobase editor does not further encode a gRNA, a separate nucleic acid vector encoding the gRNA may be administered together with the compositions described herein.

[0002] Exemplary suitable diseases and disorders include, without limitation, [[The following diseases were included in the C to T editor application. Please indicate any that are still relevant and could be treated using an adenosine deaminase.]]cystic fibrosis (see, *e.g.*, Schwank *et al.*,

Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell stem cell*. **2013**; 13: 653-658; and Wu et. al., Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell stem cell.* **2013;** 13: 659-662, neither of which uses a deaminase fusion protein to correct the genetic defect); phenylketonuria – *e.g.*, phenylalanine to serine mutation at position 835 (mouse) or 240 (human) or a homologous residue in phenylalanine hydroxylase gene (T>C mutation) – see, *e.g.*, McDonald *et al.*, *Genomics.* **1997;** 39:402-405; Bernard-Soulier syndrome (BSS) – *e.g.*, phenylalanine to serine mutation at position 55 or a homologous residue, or cysteine to arginine at residue 24 or a homologous residue in the platelet membrane glycoprotein IX (T>C mutation) – see, *e.g.*, Noris *et al.*, *British Journal of Haematology.* **1997;** 97: 312-320, and Ali *et al.*, *Hematol.* **2014;** 93: 381-384; epidermolytic hyperkeratosis (EHK) – *e.g.*, leucine to proline mutation at position 160 or 161 (if counting the initiator methionine) or a homologous residue in keratin 1 (T>C mutation) – see, *e.g.*, Chipev *et al.*, *Cell.* **1992;** 70: 821-828, see also accession number P04264 in the UNIPROT database at www[dot]uniprot[dot]org; chronic obstructive pulmonary disease (COPD) – *e.g.*, leucine to proline mutation at position 54 or 55 (if counting the initiator methionine) or a homologous residue in the processed form of α_1-antitrypsin or residue 78 in the unprocessed form or a homologous residue (T>C mutation) – see, *e.g.*, Poller *et al.*, *Genomics.* **1993;** 17: 740-743, see also accession number P01011 in the UNIPROT database; Charcot-Marie-Toot disease type 4J – *e.g.*, isoleucine to threonine mutation at position 41 or a homologous residue in FIG4 (T>C mutation) – see, *e.g.*, Lenk *et al.*, PLoS Genetics. 2011; 7: e1002104; neuroblastoma (NB) – *e.g.*, leucine to proline mutation at position 197 or a homologous residue in Caspase-9 (T>C mutation) – see, *e.g.*, Kundu *et al.*, *3 Biotech.* **2013, 3:225-234**; von Willebrand disease (vWD) – *e.g.*, cysteine to arginine mutation at position 509 or a homologous residue in the processed form of von Willebrand factor, or at position 1272 or a homologous residue in the unprocessed form of von Willebrand factor (T>C mutation) – see, *e.g.*, Lavergne *et al.*, *Br. J. Haematol.* **1992**, see also accession number P04275 in the UNIPROT database**;** 82: 66-72; myotonia congenital – *e.g.*, cysteine to arginine mutation at position 277 or a homologous residue in the muscle chloride channel gene CLCN1 (T>C mutation) – see, *e.g.*, Weinberger *et al.*, *The J. of Physiology.* **2012;** 590: 3449-3464; hereditary renal amyloidosis – *e.g.*, stop codon to arginine mutation at position 78 or a homologous residue in the processed form of apolipoprotein AII or at position 101 or a homologous residue in the unprocessed form (T>C mutation) – see, *e.g.*, Yazaki *et al.*, *Kidney*

Int. **2003;** 64: 11-16; dilated cardiomyopathy (DCM) – *e.g.*, tryptophan to Arginine mutation at position 148 or a homologous residue in the FOXD4 gene (T>C mutation), see, *e.g.*, Minoretti et. al., *Int. J. of Mol. Med.* **2007;** 19: 369-372; hereditary lymphedema – *e.g.*, histidine to arginine mutation at position 1035 or a homologous residue in VEGFR3 tyrosine kinase (A>G mutation), see, *e.g.*, Irrthum *et al.*, *Am. J. Hum. Genet.* **2000;** 67: 295-301; familial Alzheimer's disease – *e.g.*, isoleucine to valine mutation at position 143 or a homologous residue in presenilin1 (A>G mutation), see, *e.g.*, Gallo et. al., *J. Alzheimer's disease.* **2011;** 25: 425-431; Prion disease – *e.g.*, methionine to valine mutation at position 129 or a homologous residue in prion protein (A>G mutation) – see, *e.g.*, Lewis et. al., *J. of General Virology.* **2006;** 87: 2443-2449; chronic infantile neurologic cutaneous articular syndrome (CINCA) – *e.g.*, Tyrosine to Cysteine mutation at position 570 or a homologous residue in cryopyrin (A>G mutation) – see, *e.g.*, Fujisawa et. al. *Blood.* **2007;** 109: 2903-2911; and desmin-related myopathy (DRM) – *e.g.*, arginine to glycine mutation at position 120 or a homologous residue in αβ crystallin (A>G mutation) – see, *e.g.*, Kumar *et al.*, *J. Biol. Chem.* **1999;** 274: 24137-24141. The entire contents of all references and database entries is incorporated herein by reference.

[00134] Suitable routes of administrating the composition for pain suppression include, without limitation: topical, subcutaneous, transdermal, intradermal, intralesional, intraarticular, intraperitoneal, intravesical, transmucosal, gingival, intradental, intracochlear, transtympanic, intraorgan, epidural, intrathecal, intramuscular, intravenous, intravascular, intraosseus, periocular, intratumoral, intracerebral, parenteral, and intracerebroventricular administration.

[00135] The compositions of this disclosure may be administered or packaged as a unit dose, for example. The term "unit dose" when used in reference to a pharmaceutical composition of the present disclosure refers to physically discrete units suitable as unitary dosage for the subject, each unit containing a predetermined quantity of active material calculated to produce the desired therapeutic effect in association with the required diluent, *i.e.*, a carrier or vehicle.

[00136] Treatment of a disease or disorder includes delaying the development or progression of the disease, or reducing disease severity. Treating the disease does not necessarily require curative results.

[00137] As used therein, "delaying" the development of a disease means to defer, hinder, slow, retard, stabilize, and/or postpone progression of the disease. This delay can be of varying lengths of time, depending on the history of the disease and/or individuals being treated. A method that

"delays" or alleviates the development of a disease, or delays the onset of the disease, is a method that reduces probability of developing one or more symptoms of the disease in a given time frame and/or reduces extent of the symptoms in a given time frame, when compared to not using the method. Such comparisons are typically based on clinical studies, using a number of subjects sufficient to give a statistically significant result.

[00138] "Development" or "progression" of a disease means initial manifestations and/or ensuing progression of the disease. Development of the disease can be detectable and assessed using standard clinical techniques as well known in the art. However, development also refers to progression that may be undetectable. For purpose of this disclosure, development or progression refers to the biological course of the symptoms. "Development" includes occurrence, recurrence, and onset.

[00139] As used herein "onset" or "occurrence" of a disease includes initial onset and/or recurrence. Conventional methods, known to those of ordinary skill in the art of medicine, can be used to administer the isolated polypeptide or pharmaceutical composition to the subject, depending upon the type of disease to be treated or the site of the disease.

Kits

[00140] The compositions of the present disclosure may be assembled into kits. In some embodiments, the kit comprises nucleic acid vectors for the expression of the nucleobase editors described herein. In some embodiments, the kit further comprises appropriate guide nucleotide sequences (*e.g.*, gRNAs) or nucleic acid vectors for the expression of such guide nucleotide sequences, to target the Cas9 protein or nucleobase editor to the desired target sequence.

[00141] The kit described herein may include one or more containers housing components for performing the methods described herein and optionally instructions for use. Any of the kit described herein may further comprise components needed for performing the assay methods. Each component of the kits, where applicable, may be provided in liquid form (*e.g.*, in solution) or in solid form, (*e.g.*, a dry powder). In certain cases, some of the components may be reconstitutable or otherwise processible (*e.g.*, to an active form), for example, by the addition of a suitable solvent or other species (for example, water), which may or may not be provided with the kit.

[00142] In some embodiments, the kits may optionally include instructions and/or promotion for use of the components provided. As used herein, "instructions" can define a component of instruction and/or promotion, and typically involve written instructions on or associated with packaging of the disclosure. Instructions also can include any oral or electronic instructions provided in any manner such that a user will clearly recognize that the instructions are to be associated with the kit, for example, audiovisual (*e.g.*, videotape, DVD, *etc.*), Internet, and/or web-based communications, *etc.* The written instructions may be in a form prescribed by a governmental agency regulating the manufacture, use, or sale of pharmaceuticals or biological products, which can also reflect approval by the agency of manufacture, use or sale for animal administration. As used herein, "promoted" includes all methods of doing business including methods of education, hospital and other clinical instruction, scientific inquiry, drug discovery or development, academic research, pharmaceutical industry activity including pharmaceutical sales, and any advertising or other promotional activity including written, oral and electronic communication of any form, associated with the disclosure. Additionally, the kits may include other components depending on the specific application, as described herein.

[00143] The kits may contain any one or more of the components described herein in one or more containers. The components may be prepared sterilely, packaged in a syringe and shipped refrigerated. Alternatively it may be housed in a vial or other container for storage. A second container may have other components prepared sterilely. Alternatively the kits may include the active agents premixed and shipped in a vial, tube, or other container.

[00144] The kits may have a variety of forms, such as a blister pouch, a shrink wrapped pouch, a vacuum sealable pouch, a sealable thermoformed tray, or a similar pouch or tray form, with the accessories loosely packed within the pouch, one or more tubes, containers, a box or a bag. The kits may be sterilized after the accessories are added, thereby allowing the individual accessories in the container to be otherwise unwrapped. The kits can be sterilized using any appropriate sterilization techniques, such as radiation sterilization, heat sterilization, or other sterilization methods known in the art. The kits may also include other components, depending on the specific application, for example, containers, cell media, salts, buffers, reagents, syringes, needles, a fabric, such as gauze, for applying or removing a disinfecting agent, disposable gloves, a support for the agents prior to administration, *etc.*

Host Cells

[00145] Cells that may contain any of the compositions described herein include prokaryotic cells and eukaryotic cells. The methods described herein are used to deliver a Cas9 protein or a nucleobase editor into a eukaryotic cell (*e.g.*, a mammalian cell, such as a human cell). In some embodiments, the cell is in vitro (*e.g.*, cultured cell. In some embodiments, the cell is in vivo (*e.g.*, in a subject such as a human subject). In some embodiments, the cell is ex vivo (*e.g.*, isolated from a subject and may be administered back to the same or a different subject).

[00146] Mammalian cells of the present disclosure include human cells, primate cells (*e.g.*, vero cells), rat cells (*e.g.*, GH3 cells, OC23 cells) or mouse cells (*e.g.*, MC3T3 cells). There are a variety of human cell lines, including, without limitation, human embryonic kidney (HEK) cells, HeLa cells, cancer cells from the National Cancer Institute's 60 cancer cell lines (NCI60), DU145 (prostate cancer) cells, Lncap (prostate cancer) cells, MCF-7 (breast cancer) cells, MDA-MB-438 (breast cancer) cells, PC3 (prostate cancer) cells, T47D (breast cancer) cells, THP-1 (acute myeloid leukemia) cells, U87 (glioblastoma) cells, SHSY5Y human neuroblastoma cells (cloned from a myeloma) and Saos-2 (bone cancer) cells. In some embodiments, rAAV vectors are delivered into human embryonic kidney (HEK) cells (*e.g.*, HEK 293 or HEK 293T cells). In some embodiments, rAAV vectors are delivered into stem cells (*e.g.*, human stem cells) such as, for example, pluripotent stem cells (*e.g.*, human pluripotent stem cells including human induced pluripotent stem cells (hiPSCs)). A stem cell refers to a cell with the ability to divide for indefinite periods in culture and to give rise to specialized cells. A pluripotent stem cell refers to a type of stem cell that is capable of differentiating into all tissues of an organism, but not alone capable of sustaining full organismal development. A human induced pluripotent stem cell refers to a somatic (*e.g.*, mature or adult) cell that has been reprogrammed to an embryonic stem cell-like state by being forced to express genes and factors important for maintaining the defining properties of embryonic stem cells (*see*, *e.g.*, Takahashi and Yamanaka, *Cell* 126 (4): 663–76, 2006, incorporated by reference herein). Human induced pluripotent stem cell cells express stem cell markers and are capable of generating cells characteristic of all three germ layers (ectoderm, endoderm, mesoderm).

[00147] Additional non-limiting examples of cell lines that may be used in accordance with the present disclosure include 293-T, 293-T, 3T3, 4T1, 721, 9L, A-549, A172, A20, A253, A2780, A2780ADR, A2780cis, A431, ALC, B16, B35, BCP-1, BEAS-2B, bEnd.3, BHK-21, BR 293,

BxPC3, C2C12, C3H-10T1/2, C6, C6/36, Cal-27, CGR8, CHO, CML T1, CMT, COR-L23, COR-L23/5010, COR-L23/CPR, COR-L23/R23, COS-7, COV-434, CT26, D17, DH82, DU145, DuCaP, E14Tg2a, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, Hepa1c1c7, High Five cells, HL-60, HMEC, HT-29, HUVEC, J558L cells, Jurkat, JY cells, K562 cells, KCL22, KG1, Ku812, KYO1, LNCap, Ma-Mel 1, 2, 3....48, MC-38, MCF-10A, MCF-7, MDA-MB-231, MDA-MB-435, MDA-MB-468, MDCK II, MG63, MONO-MAC 6, MOR/0.2R, MRC5, MTD-1A, MyEnd, NALM-1, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NW-145, OPCN/OPCT Peer, PNT-1A/PNT 2, PTK2, Raji, RBL cells, RenCa, RIN-5F, RMA/RMAS, S2, Saos-2 cells, Sf21, Sf9, SiHa, SKBR3, SKOV-3, T-47D, T2, T84, THP1, U373, U87, U937, VCaP, WM39, WT-49, X63, YAC-1 and YAR cells.

[00148] Without further elaboration, it is believed that one skilled in the art can, based on the above description, utilize the present disclosure to its fullest extent. The following specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever. All publications cited herein are incorporated by reference for the purposes or subject matter referenced herein.

EXAMPLES

[00149] In order that the invention described herein may be more fully understood, the following examples are set forth. The synthetic examples described in this application are offered to illustrate the compounds and methods provided herein and are not to be construed in any way as limiting their scope.

Example 1: Amino Acid Sequences of Cas9 Proteins and Nucleobase Editor

[00150] Non-limiting examples of suitable Cas9 proteins and variants, and nucleobase editors and variants are provided. The disclosure provides Cas9 variants, for example, Cas9 proteins from one or more organisms, which may comprise one or more mutations (*e.g.*, to generate dCas9 or Cas9 nickase). In some embodiments, one or more of the amino acid residues, identified below by an asterisk, of a Cas9 protein may be mutated. In some embodiments, the D10 and/or H840 residues of the amino acid sequence provided in SEQ ID NO: 1, or a corresponding mutation in any of the amino acid sequences provided in SEQ ID NOs: 2-275 and

394-397, are mutated. In some embodiments, the D10 residue of the amino acid sequence provided in SEQ ID NO: 1, or a corresponding mutation in any of the amino acid sequences provided in SEQ ID NOs: 2-275 and 394-397, is mutated to any amino acid residue, except for D. In some embodiments, the D10 residue of the amino acid sequence provided in SEQ ID NO: 1, or a corresponding mutation in any of the amino acid sequences provided in SEQ ID NOs: 2-275 and 394-397, is mutated to an A. In some embodiments, the H840 residue of the amino acid sequence provided in SEQ ID NO: 1, or a corresponding residue in any of the amino acid sequences provided in SEQ ID NOs: 2-275 and 394-397, is an H. In some embodiments, the H840 residue of the amino acid sequence provided in SEQ ID NO: 1, or a corresponding mutation in any of the amino acid sequences provided in SEQ ID NOs: 2-275 and 394-397, is mutated to any amino acid residue, except for H. In some embodiments, the H840 residue of the amino acid sequence provided in SEQ ID NO: 1, or a corresponding mutation in any of the amino acid sequences provided in SEQ ID NOs: 2-275 and 394-397, is mutated to an A. In some embodiments, the D10 residue of the amino acid sequence provided in SEQ ID NO: 1, or a corresponding residue in any of the amino acid sequences provided in SEQ ID NOs: 2-275 and 394-397, is a D.

[00151] A number of Cas9 sequences from various species were aligned to determine whether corresponding homologous amino acid residues of D10 and H840 of SEQ ID NO: 1 can be identified in other Cas9 proteins, allowing the generation of Cas9 variants with corresponding mutations of the homologous amino acid residues. The alignment was carried out using the NCBI Constraint-based Multiple Alignment Tool (COBALT (accessible at st-va.ncbi.nlm.nih.gov/tools/cobalt)), with the following parameters. Alignment parameters: Gap penalties -11,-1; End-Gap penalties -5,-1. CDD Parameters: Use RPS BLAST on; Blast E-value 0.003; Find Conserved columns and Recompute on. Query Clustering Parameters: Use query clusters on; Word Size 4; Max cluster distance 0.8; Alphabet Regular.

S. pyogenes Cas9 wild type (NCBI Reference Sequence: NC_002737.2, Uniprot Reference Sequence: Q99ZW2)

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR
YTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV
DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSAR
LSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQY
ADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG

YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLP
NEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE
CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDK
VMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGI
KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTR
SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITK
HVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIK
KYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI
VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS
VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADAN
LDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD (SEQ ID NO: 1)

S. pyogenes dCas9 (D10A and H840A)

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR
YTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV
DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSAR
LSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQY
ADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG
YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLP
NEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE
CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDK
VMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGI
KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTR
SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITK
HVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIK
KYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI
VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS
VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADAN
LDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD (SEQ ID NO: 2)

S. pyogenes Cas9 Nickase (D10A)

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR
YTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV
DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSAR
LSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQY
ADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG
YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLP
NEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE
CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDK
VMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGI
KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTR
SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITK

HVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIK
KYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI
VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS
VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADAN
LDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD (SEQ ID NO: 3)

VRER-nCas9 (D10A/D1135V/G1218R/R1335E/T1337R) *S. pyogenes* Cas9 Nickase

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR
YTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV
DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSAR
LSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQY
ADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG
YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLP
NEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE
CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDK
VMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGI
KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTR
SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITK
HVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIK
KYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI
VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS
VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADAN
LDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKEYRSTKEVLDATLIHQSITGLYETRI
DLSQLGGD (SEQ ID NO: 4)

VQR-nCas9 (D10A/D1135V/R1335Q/T1337R) *S. pyogenes* Cas9 Nickase

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR
YTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV
DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSAR
LSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQY
ADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG
YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLP
NEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE
CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDK
VMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGI
KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTR
SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITK
HVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIK
KYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI
VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS
VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADAN
LDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETR
IDLSQLGGD (SEQ ID NO: 5)

EQR-nCas9 (D10A/D1135E/R1335Q/T1337R) *S. pyogenes* Cas9 Nickase

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR
YTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV
DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSAR
LSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQY
ADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG
YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLP
NEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE
CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDK
VMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGI
KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTR
SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITK
HVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIK
KYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI
VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFESPTVAYS
VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADAN
LDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETR
IDLSQLGGD (SEQ ID NO: 6)

KKH-nCas9 (D10A/E782K/N968K/R1015H) *S. aureus* Cas9 Nickase

MKRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLF
DYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSK
ALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEG
PGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQII
ENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIY
QSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQ
QKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEE
IIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEEN
SKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYAT
RGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDK
AKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKD
DKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGN
YLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLD
VIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYKNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYL
ENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG (SEQ ID NO: 7)

Streptococcus thermophilus CRISPR1 Cas9 (St1Cas9) Nickase (D9A)

MSDLVLGLAIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLTRRKKHRRVRLNRLFEE
SGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSIGDYAQIVKENSKQ
LETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEI
LTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTV
PTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLE
TLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGWHNFSVKL
MMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFD
NIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQG
ERCLYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRE
LKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRG
QFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVF

KAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYTQDGYD
AFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGP
EIKSLKYYDSKLGNHIDITPKDSNNKVVLQSVSPWRADVYFNKTTGKYEILGLKYADLQFEKGTGTYKISQ
EKYNDIKKKEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTMPKQKHYVELKPYDKQKFEGGEA
LIKVLGNVANSGQCKKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPKLDF (SEQ ID NO: 8)

Streptococcus thermophilus CRISPR3Cas9 (St3Cas9) Nickase (D10A)

MTKPYSIGLAIGTNSVGWAVITDNYKVPSKKMKVLGNTSKKYIKKNLLGVLLFDSGITAEGRRLKRTARRR
YTRRRNRILYLQEIFSTEMATLDDAFFQRLDDSFLVPDDKRDSKYPIFGNLVEEKVYHDEFPTIYHLRKYLA
DSTKKADLRLVYLALAHMIKYRGHFLIEGEFNSKNNDIQKNFQDFLDTYNAIFESDLSLENSKQLEEIVKDK
ISKLEKKDRILKLFPGEKNSGIFSEFLKLIVGNQADFRKCFNLDEKASLHFSKESYDEDLETLLGYIGDDYSD
VFLKAKKLYDAILLSGFLTVTDNETEAPLSSAMIKRYNEHKEDLALLKEYIRNISLKTYNEVFKDDTKNGY
AGYIDGKTNQEDFYVYLKNLLAEFEGADYFLEKIDREDFLRKQRTFDNGSIPYQIHLQEMRAILDKQAKFY
PFLAKNKERIEKILTFRIPYYVGPLARGNSDFAWSIRKRNEKITPWNFEDVIDKESSAEAFINRMTSFDLYLPE
EKVLPKHSLLYETFNVYNELTKVRFIAESMRDYQFLDSKQKKDIVRLYFKDKRKVTDKDIIEYLHAIYGYD
GIELKGIEKQFNSSLSTYHDLLNIINDKEFLDDSSNEAIIEEIIHTLTIFEDREMIKQRLSKFENIFDKSVLKKLS
RRHYTGWGKLSAKLINGIRDEKSGNTILDYLIDDGISNRNFMQLIHDDALSFKKKIQKAQIIGDEDKGNIKEV
VKSLPGSPAIKKGILQSIKIVDELVKVMGGRKPESIVVEMARENQYTNQGKSNSQQRLKRLEKSLKELGSKI
LKENIPAKLSKIDNNALQNDRLYLYYLQNGKDMYTGDDLDIDRLSNYDIDHIIPQAFLKDNSIDNKVLVSSA
SNRGKSDDFPSLEVVKKRKTFWYQLLKSKLISQRKFDNLTKAERGGLLPEDKAGFIQRQLVETRQITKHVA
RLLDEKFNNKKDENNRAVRTVKIITLKSTLVSQFRKDFELYKVREINDFHHAHDAYLNAVIASALLKKYPK
LEPEFVYGDYPKYNSFRERKSATEKVYFYSNIMNIFKKSISLADGRVIERPLIEVNEETGESVWNKESDLATV
RRVLSYPQVNVVKKVEEQNHGLDRGKPKGLFNANLSSKPKPNSNENLVGAKEYLDPKKYGGYAGISNSFA
VLVKGTIEKGAKKKITNVLEFQGISILDRINYRKDKLNFLLEKGYKDIELIIELPKYSLFELSDGSRRMLASILS
TNNKRGEIHKGNQIFLSQKFVKLLYHAKRISNTINENHRKYVENHKKEFEELFYYILEFNENYVGAKKNGK
LLNSAFQSWQNHSIDELCSSFIGPTGSERKGLFELTSRGSAADFEFLGVKIPRYRDYTPSSLLKDATLIHQSVT
GLYETRIDLAKLGEG (SEQ ID NO: 9)

S. aureus Cas9 wild type

MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLF
DYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSK
ALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEG
PGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQII
ENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIY
QSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQ
QKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEE
IIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEEN
SKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYAT
RGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDK
AKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDD
KGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNY
LTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVI
KKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLE
NMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG (SEQ ID NO: 10)

S. aureus Cas9 Nickase (D10A)

MKRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLF
DYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSK
ALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEG
PGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQII

ENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIY
QSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQ
QKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEE
IIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEEN
SKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYAT
RGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDK
AKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDD
KGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNY
LTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVI
KKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLE
NMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG (SEQ ID NO: 11)

Streptococcus thermophilus wild type CRISPR3 Cas9 (St3Cas9)

MTKPYSIGLDIGTNSVGWAVITDNYKVPSKKMKVLGNTSKKYIKKNLLGVLLFDSGITAEGRRLKRTARRR
YTRRRNRILYLQEIFSTEMATLDDAFFQRLDDSFLVPDDKRDSKYPIFGNLVEEKVYHDEFPTIYHLRKYLA
DSTKKADLRLVYLALAHMIKYRGHFLIEGEFNSKNNDIQKNFQDFLDTYNAIFESDLSLENSKQLEEIVKDK
ISKLEKKDRILKLFPGEKNSGIFSEFLKLIVGNQADFRKCFNLDEKASLHFSKESYDEDLETLLGYIGDDYSD
VFLKAKKLYDAILLSGFLTVTDNETEAPLSSAMIKRYNEHKEDLALLKEYIRNISLKTYNEVFKDDTKNGY
AGYIDGKTNQEDFYVYLKNLLAEFEGADYFLEKIDREDFLRKQRTFDNGSIPYQIHLQEMRAILDKQAKFY
PFLAKNKERIEKILTFRIPYYVGPLARGNSDFAWSIRKRNEKITPWNFEDVIDKESSAEAFINRMTSFDLYLPE
EKVLPKHSLLYETFNVYNELTKVRFIAESMRDYQFLDSKQKKDIVRLYFKDKRKVTDKDIIEYLHAIYGYD
GIELKGIEKQFNSSLSTYHDLLNIINDKEFLDDSSNEAIIEEIIHTLTIFEDREMIKQRLSKFENIFDKSVLKKLS
RRHYTGWGKLSAKLINGIRDEKSGNTILDYLIDDGISNRNFMQLIHDDALSFKKKIQKAQIIGDEDKGNIKEV
VKSLPGSPAIKKGILQSIKIVDELVKVMGGRKPESIVVEMARENQYTNQGKSNSQQRLKRLEKSLKELGSKI
LKENIPAKLSKIDNNALQNDRLYLYYLQNGKDMYTGDDLDIDRLSNYDIDHIIPQAFLKDNSIDNKVLVSSA
SNRGKSDDFPSLEVVKKRKTFWYQLLKSKLISQRKFDNLTKAERGGLLPEDKAGFIQRQLVETRQITKHVA
RLLDEKFNNKKDENNRAVRTVKIITLKSTLVSQFRKDFELYKVREINDFHHAHDAYLNAVIASALLKKYPK
LEPEFVYGDYPKYNSFRERKSATEKVYFYSNIMNIFKKSISLADGRVIERPLIEVNEETGESVWNKESDLATV
RRVLSYPQVNVVKKVEEQNHGLDRGKPKGLFNANLSSKPKPNSNENLVGAKEYLDPKKYGGYAGISNSFA
VLVKGTIEKGAKKKITNVLEFQGISILDRINYRKDKLNFLLEKGYKDIELIIELPKYSLFELSDGSRRMLASILS
TNNKRGEIHKGNQIFLSQKFVKLLYHAKRISNTINENHRKYVENHKKEFEELFYYILEFNENYVGAKKNGK
LLNSAFQSWQNHSIDELCSSFIGPTGSERKGLFELTSRGSAADFEFLGVKIPRYRDYTPSSLLKDATLIHQSVT
GLYETRIDLAKLGEG (SEQ ID NO: 12)

Streptococcus thermophilus CRISPR1 Cas9 wild type (St1Cas9)

MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLTRRKKHRRVRLNRLFEE
SGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSIGDYAQIVKENSKQ
LETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEI
LTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTV
PTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLE
TLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGWHNFSVKL
MMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFD
NIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQG
ERCLYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRE
LKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRG
QFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVF
KAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYTQDGYD
AFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGP
EIKSLKYYDSKLGNHIDITPKDSNNKVVLQSVSPWRADVYFNKTTGKYEILGLKYADLQFEKGTGTYKISQ
EKYNDIKKKEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTMPKQKHYVELKPYDKQKFEGGEA
LIKVLGNVANSGQCKKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPKLDF (SEQ ID NO: 13)

CasX from *Sulfolobus islandicus* (strain REY15A).

MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGKAKKKKGEEG
ETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVKPEFYKFGRSPGMVERT
RRVKLEVEPHYLIMAAAGWVLTRLGKAKVSEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWI
ARKVVSSVTNPNVSVVSIYTISDAVGQNPTTINGGFSIDLTKLLEKRDLLSERLEAIARNALSISSNMRERYIV
LANYIYEYLTGSKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG (SEQ ID NO: 14)

CasY from *Sulfolobus islandicus* (strain REY15A).

MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGKAKKKKGEEG
ETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVKPEFYEFGRSPGMVERT
RRVKLEVEPHYLIIAAAGWVLTRLGKAKVSEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIA
RKVVSSVTNPNVSVVRIYTISDAVGQNPTTINGGFSIDLTKLLEKRYLLSERLEAIARNALSISSNMRERYIVL
ANYIYEYLTGSKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG (SEQ ID NO: 15)

Wild type *Francisella novicida* Cpf1 (D917, E1006, and D1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVCISED
LLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQS
KDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYE
SLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNG
ENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKT
VEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKE
QELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQG
KKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKI
RNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGE
GYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYK
QSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSK
GRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIK
DKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNI
IGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFK
RGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTS
KICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFR
NSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTEL
DYLISPVADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQ
NRNN (SEQ ID NO: 16)

Francisella novicida Cpf1 D917A (A917, E1006, and D1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVCISED
LLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQS
KDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYE
SLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNG
ENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKT
VEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKE
QELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQG
KKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKI
RNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGE
GYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYK
QSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSK
GRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIK
DKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNI
IGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFK

RGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTS
KICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFR
NSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTEL
DYLISPVADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQ
NRNN (SEQ ID NO: 17)

Francisella novicida Cpf1 E1006A (D917, A1006, and D1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVCISED
LLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQS
KDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYE
SLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNG
ENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKT
VEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKE
QELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQG
KKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKI
RNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGE
GYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYK
QSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSK
GRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIK
DKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNI
IGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**A**DLNFGFK
RGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTS
KICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFR
NSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTEL
DYLISPVADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQ
NRNN (SEQ ID NO: 18)

Francisella novicida Cpf1 D1255A (D917, E1006, and A1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVCISED
LLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQS
KDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYE
SLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNG
ENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKT
VEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKE
QELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQG
KKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKI
RNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGE
GYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYK
QSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSK
GRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIK
DKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNI
IGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFK
RGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTS
KICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFR
NSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTEL
DYLISPVADVNGNFFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQ
NRNN (SEQ ID NO: 19)

Francisella novicida Cpf1 D917A/E1006A (A917, A1006, and D1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVCISED
LLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQS
KDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYE
SLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNG
ENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKT
VEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKE
QELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQG
KKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKI
RNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGE
GYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYK
QSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSK
GRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIK
DKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNI
IGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**A**DLNFGFK
RGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTS
KICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFR
NSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTEL
DYLISPVADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQ
NRNN (SEQ ID NO: 20)

Francisella novicida Cpf1 D917A/D1255A (A917, E1006, and A1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVCISED
LLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQS
KDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYE
SLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNG
ENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKT
VEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKE
QELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQG
KKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKI
RNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGE
GYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYK
QSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSK
GRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIK
DKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNI
IGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFK
RGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTS
KICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFR
NSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTEL
DYLISPVADVNGNFFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQ
NRNN (SEQ ID NO: 21)

Francisella novicida Cpf1 E1006A/D1255A (D917, A1006, and A1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVCISED
LLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQS
KDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYE

SLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNG
ENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKT
VEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKE
QELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQG
KKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKI
RNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGE
GYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYK
QSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSK
GRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIK
DKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTFNI
IGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFADLNFGFK
RGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTS
KICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFR
NSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTEL
DYLISPVADVNGNFFDSRQAPKNMPQDAAANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQ
NRNN (SEQ ID NO: 22)

Francisella novicida Cpf1 D917A/E1006A/D1255A (A917, A1006, and A1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVCISED
LLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQS
KDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYE
SLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNG
ENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKT
VEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKE
QELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQG
KKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKI
RNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGE
GYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYK
QSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSK
GRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIK
DKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIARGERHLAYYTLVDGKGNIIKQDTFNI
IGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFADLNFGFK
RGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTS
KICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFR
NSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTEL
DYLISPVADVNGNFFDSRQAPKNMPQDAAANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQ
NRNN (SEQ ID NO: 23)

Wild type *Natronobacterium gregoryi* Argonaute

MTVIDLDSTTTADELTSGHTYDISVTLTGVYDNTDEQHPRMSLAFEQDNGERRYITLWKNTTPKDVFTYDY
ATGSTYIFTNIDYEVKDGYENLTATYQTTVENATAQEVGTTDEDETFAGGEPLDHHLDDALNETPDDAETE
SDSGHVMTSFASRDQLPEWTLHTYTLTATDGAKTDTEYARRTLAYTVRQELYTDHDAAPVATDGLMLLT
PEPLGETPLDLDCGVRVEADETRTLDYTTAKDRLLARELVEEGLKRSLWDDYLVRGIDEVLSKEPVLTCDE
FDLHERYDLSVEVGHSGRAYLHINFRHRFVPKLTLADIDDDNIYPGLRVKTTYRPRRGHIVWGLRDECATD
SLNTLGNQSVVAYHRNNQTPINTDLLDAIEAADRRVVETRRQGHGDDAVSFPQELLAVEPNTHQIKQFASD
GFHQQARSKTRLSASRCSEKAQAFAERLDPVRLNGSTVEFSSEFFTGNNEQQLRLLYENGESVLTFRDGAR
GAHPDETFSKGIVNPPESFEVAVVLPEQQADTCKAQWDTMADLLNQAGAPPTRSETVQYDAFSSPESISLN
VAGAIDPSEVDAAFVVLPPDQEGFADLASPTETYDELKKALANMGIYSQMAYFDRFRDAKIFYTRNVALG
LLAAAGGVAFTTEHAMPGDADMFIGIDVSRSYPEDGASGQINIAATATAVYKDGTILGHSSTRPQLGEKLQ
STDVRDIMKNAILGYQQVTGESPTHIVIHRDGFMNEDLDPATEFLNEQGVEYDIVEIRKQPQTRLLAVSDVQ

YDTPVKSIAAINQNEPRATVATFGAPEYLATRDGGGLPRPIQIERVAGETDIETLTRQVYLLSQSHIQVHNST
ARLPITTAYADQASTHATKGYLVQTGAFESNVGFL (SEQ ID NO: 24)

Cas9 variant with decreased electrostatic interactions between the Cas9 and DNA backbone

DKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRY
TRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLS
KSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYAD
LFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFL
KDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT**A**FDKNLPNE
KVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVM
KQLKRRRYTGWG**A**LSRKLINGIRDKQSGKTILDFLKSDGFANRNFM**A**LIHDDSLTFKEDIQKAQVSGQGDS
LHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE
LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSD
KNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR**A**ITKHV
AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYP
KLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVW
DKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL
VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLD
KVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDL
SQLGGD (SEQ ID NO: 25)CasY (ncbi.nlm.nih.gov/protein/APG80656.1)

>APG80656.1 CRISPR-associated protein CasY [uncultured *Parcubacteria* group bacterium]

MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIVSAINDDYVGLYGLSNFDDL
YNAEKRNEEKVYSVLDFWYDCVQYGAVFSYTAPGLLKNVAEVRGGSYELTKTLKGSHLYDELQIDKVIKF
LNKKEISRANGSLDKLKKDIIDCFKAEYRERHKDQCNKLADDIKNAKKDAGASLGERQKKLFRDFFGISEQ
SENDKPSFTNPLNLTCCLLPFDTVNNNRNRGEVLFNKLKEYAQKLDKNEGSLEMWEYIGIGNSGTAFSNFL
GEGFLGRLRENKITELKKAMMDITDAWRGQEQEEELEKRLRILAALTIKLREPKFDNHWGGYRSDINGKLS
SWLQNYINQTVKIKEDLKGHKKDLKKAKEMINRFGESDTKEEAVVSSLLESIEKIVPDDSADDEKPDIPAIAI
YRRFLSDGRLTLNRFVQREDVQEALIKERLEAEKKKKPKKRKKKSDAEDEKETIDFKELFPHLAKPLKLVP
NFYGDSKRELYKKYKNAAIYTDALWKAVEKIYKSAFSSSLKNSFFDTDFDKDFFIKRLQKIFSVYRRFNTDK
WKPIVKNSFAPYCDIVSLAENEVLYKPKQSRSRKSAAIDKNRVRLPSTENIAKAGIALARELSVAGFDWKDL
LKKEEHEEYIDLIELHKTALALLLAVTETQLDISALDFVENGTVKDFMKTRDGNLVLEGRFLEMFSQSIVFS
ELRGLAGLMSRKEFITRSAIQTMNGKQAELLYIPHEFQSAKITTPKEMSRAFLDLAPAEFATSLEPESLSEKS
LLKLKQMRYYPHYFGYELTRTGQGIDGGVAENALRLEKSPVKKREIKCKQYKTLGRGQNKIVLYVRSSYY
QTQFLEWFLHRPKNVQTDVAVSGSFLIDEKKVKTRWNYDALTVALEPVSGSERVFVSQPFTIFPEKSAEEE
GQRYLGIDIGEYGIAYTALEITGDSAKILDQNFISDPQLKTLREEVKGLKLDQRRGTFAMPSTKIARIRESLV
HSLRNRIHHLALKHKAKIVYELEVSRFEEGKQKIKKVYATLKKADVYSEIDADKNLQTTVWGKLAVASEIS
ASYTSQFCGACKKLWRAEMQVDETITTQELIGTVRVIKGGTLIDAIKDFMRPPIFDENDTPFPKYRDFCDKH
HISKKMRGNSCLFICPFCRANADADIQASQTIALLRYVKEEKKVEDYFERFRKLKNIKVLGQMKKI (SEQ
ID NO: 26)

High-fidelity Cas9 domain

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRY
TRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLS
KSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYAD
LFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFL
KDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTAFDKNLPNE
KVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVM
KQLKRRRYTGWGALSRKLINGIRDKQSGKTILDFLKSDGFANRNFMALIHDDSLTFKEDIQKAQVSGQGDS
LHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE
LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSD
KNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRAITKHV
AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYP
KLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVW
DKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL
VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLD
KVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDL
SQLGGD (SEQ ID NO: 394)

C2c1 (uniprot.org/uniprot/T0D7A2#)

sp|T0D7A2|C2C1_ALIAG CRISPR-associated endonuclease C2c1 OS=Alicyclobacillus acidoterrestris (strain ATCC 49025 / DSM 3922 / CIP 106132 / NCIMB 13137 / GD3B) GN=c2c1 PE=1 SV=1

MAVKSIKVKLRLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQECDKTAEECK
AELLERLRARQVENGHRGPAGSDDELLQLARQLYELLVPQAIGAKGDAQQIARKFLSPLADKDAVGGLGI
AKAGNKPRWVRMREAGEPGWEEEKEKAETRKSADRTADVLRALADFGLKPLMRVYTDSEMSSVEWKPL
RKGQAVRTWDRDMFQQAIERMMSWESWNQRVGQEYAKLVEQKNRFEQKNFVGQEHLVHLVNQLQQD
MKEASPGLESKEQTAHYVTGRALRGSDKVFEKWGKLAPDAPFDLYDAEIKNVQRRNTRRFGSHDLFAKL
AEPEYQALWREDASFLTRYAVYNSILRKLNHAKMFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGE
RRHAIRFHKLLKVENGVAREVDDVTVPISMSEQLDNLLPRDPNEPIALYFRDYGAEQHFTGEFGGAKIQCR
RDQLAHMHRRRGARDVYLNVSVRVQSQSEARGERRPPYAAVFRLVGDNHRAFVHFDKLSDYLAEHPDD
GKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSKGRVPFFFPIKGNDNLVAVHERSQLLKLPG
ETESKDLRAIREERQRTLRQLRTQLAYLRLLVRCGSEDVGRRERSWAKLIEQPVDAANHMTPDWREAFEN
ELQKLKSLHGICSDKEWMDAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYAKDVVGGNSIEQIEY
LERQYKFLKSWSFFGKVSGQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRIIMEALGYVYALDERGKGK
WVAKYPPCQLILLEELSEYQFNNDRPPSENNQLMQWSHRGVFQELINQAQVHDLLVGTMYAAFSSRFDAR
TGAPGIRCRRVPARCTQEHNPEPFPWWLNKFVVEHTLDACPLRADDLIPTGEGEIFVSPFSAEEGDFHQIHA
DLNAAQNLQQRLWSDFDISQIRLRCDWGEVDGELVLIPRLTGKRTADSYSNKVFYTNTGVTYYERERGKK
RRKVFAQEKLSEEEAELLVEADEAREKSVVLMRDPSGIINRGNWTRQKEFWSMVNQRIEGYLVKQIRSRVP
LQDSACENTGDI (SEQ ID NO: 395)

C2c2 (uniprot.org/uniprot/P0DOC6)

>sp|P0DOC6|C2C2_LEPSD CRISPR-associated endoribonuclease C2c2 OS=Leptotrichia shahii (strain DSM 19757 / CCUG 47503 / CIP 107916 / JCM 16776 / LB37) GN=c2c2 PE=1 SV=1

MGNLFGHKRWYEVRDKKDFKIKRKVKVKRNYDGNKYILNINENNNKEKIDNNKFIRKYINYKKNDNILKE
FTRKFHAGNILFKLKGKEGIIRIENNDDFLETEEVVLYIEAYGKSEKLKALGITKKKIIDEAIRQGITKDDKKIE
IKRQENEEEIEIDIRDEYTNKTLNDCSIILRIIENDELETKKSIYEIFKNINMSLYKIIEKIIENETEKVFENRYYEE
HLREKLLKDDKIDVILTNFMEIREKIKSNLEILGFVKFYLNVGGDKKKSKNKKMLVEKILNINVDLTVEDIA
DFVIKELEFWNITKRIEKVKKVNNEFLEKRRNRTYIKSYVLLDKHEKFKIERENKKDKIVKFFVENIKNNSIK
EKIEKILAEFKIDELIKKLEKELKKGNCDTEIFGIFKKHYKVNFDSKKFSKKSDEEKELYKIIYRYLKGRIEKIL
VNEQKVRLKKMEKIEIEKILNESILSEKILKRVKQYTLEHIMYLGKLRHNDIDMTTVNTDDFSRLIAKEELD
LELITFFASTNMELNKIFSRENINNDENIDFFGGDREKNYVLDKKILNSKIKIIRDLDFIDNKNNITNNFIRKFT
KIGTNERNRILHAISKERDLQGTQDDYNKVINIIQNLKISDEEVSKALNLDVVFKDKKNIITKINDIKISEENN
NDIKYLPSFSKVLPEILNLYRNNPKNEPFDTIETEKIVLNALIYVNKELYKKLILEDDLEENESKNIFLQELKK
TLGNIDEIDENIIENYYKNAQISASKGNNKAIKKYQKKVIECYIGYLRKNYEELFDFSDFKMNIQEIKKQIKDI
NDNKTYERITVKTSDKTIVINDDFEYIISIFALLNSNAVINKIRNRFFATSVWLNTSEYQNIIDILDEIMQLNTL
RNECITENWNLNLEEFIQKMKEIEKDFDDFKIQTKKEIFNNYYEDIKNNILTEFKDDINGCDVLEKKLEKIVIF
DDETKFEIDKKSNILQDEQRKLSNINKKDLKKKVDQYIKDKDQEIKSKILCRIIFNSDFLKKYKKEIDNLIED
MESENENKFQEIYYPKERKNELYIYKKNLFLNIGNPNFDKIYGLISNDIKMADAKFLFNIDGKNIRKNKISEID
AILKNLNDKLNGYSKEYKEKYIKKLKENDDFFAKNIQNKNYKSFEKDYNRVSEYKKIRDLVEFNYLNKIES
YLIDINWKLAIQMARFERDMHYIVNGLRELGIIKLSGYNTGISRAYPKRNGSDGFYTTTAYYKFFDEESYKK
FEKICYGFGIDLSENSEINKPENESIRNYISHFYIVRNPFADYSIAEQIDRVSNLLSYSTRYNNSTYASVFEVFK
KDVNLDYDELKKKFKLIGNNDILERLMKPKKVSVLELESYNSDYIKNLIIELLTKIENTNDTL (SEQ ID NO: 396)

C2c3, translated from >CEPX01008730.1 marine metagenome genome assembly TARA_037_MES_0.1-0.22, contig TARA_037_MES_0.1-0.22_scaffold22115_1, whole genome shotgun sequence.

MRSNYHGGRNARQWRKQISGLARRTKETVFTYKFPLETDAAEIDFDKAVQTYGIAEGVGHGSLIGLVCAF
HLSGFRLFSKAGEAMAFRNRSRYPTDAFAEKLSAIMGIQLPTLSPEGLDLIFQSPPRSRDGIAPVWSENEVRN
RLYTNWTGRGPANKPDEHLLEIAGEIAKQVFPKFGGWDDLASDPDKALAAADKYFQSQGDFPSIASLPAAI
MLSPANSTVDFEGDYIAIDPAAETLLHQAVSRCAARLGRERPDLDQNKGPFVSSLQDALVSSQNNGLSWLF
GVGFQHWKEKSPKELIDEYKVPADQHGAVTQVKSFVDAIPLNPLFDTTHYGEFRASVAGKVRSWVANYW
KRLLDLKSLLATTEFTLPESISDPKAVSLFSGLLVDPQGLKKVADSLPARLVSAEEAIDRLMGVGIPTAADIA
QVERVADEIGAFIGQVQQFNNQVKQKLENLQDADDEEFLKGLKIELPSGDKEPPAINRISGGAPDAAAEISE
LEEKLQRLLDARSEHFQTISEWAEENAVTLDPIAAMVELERLRLAERGATGDPEEYALRLLLQRIGRLANR
VSPVSAGSIRELLKPVFMEEREFNLFFHNRLGSLYRSPYSTSRHQPFSIDVGKAKAIDWIAGLDQISSDIEKAL
SGAGEALGDQLRDWINLAGFAISQRLRGLPDTVPNALAQVRCPDDVRIPPLLAMLLEEDDIARDVCLKAFN
LYVSAINGCLFGALREGFIVRTRFQRIGTDQIHYVPKDKAWEYPDRLNTAKGPINAAVSSDWIEKDGAVIKP
VETVRNLSSTGFAGAGVSEYLVQAPHDWYTPLDLRDVAHLVTGLPVEKNITKLKRLTNRTAFRMVGASSF
KTHLDSVLLSDKIKLGDFTIIIDQHYRQSVTYGGKVKISYEPERLQVEAAVPVVDTRDRTVPEPDTLFDHIVA
IDLGERSVGFAVFDIKSCLRTGEVKPIHDNNGNPVVGTVAVPSIRRLMKAVRSHRRRRQPNQKVNQTYSTA
LQNYRENVIGDVCNRIDTLMERYNAFPVLEFQIKNFQAGAKQLEIVYGS (SEQ ID NO: 397)

[00152] An exemplary alignment of four Cas9 sequences is provided below. The Cas9 sequences in the alignment are: Sequence 1 (S1): SEQ ID NO: 1 | WP_010922251| gi 499224711 | type II CRISPR RNA-guided endonuclease Cas9 [*Streptococcus pyogenes*]; Sequence 2 (S2): SEQ ID NO: 27 | WP_039695303 | gi 746743737 | type II CRISPR RNA-guided endonuclease Cas9 [*Streptococcus gallolyticus*]; Sequence 3 (S3): SEQ ID NO: 28 | WP_045635197 | gi 782887988 | type II CRISPR RNA-guided endonuclease Cas9 [*Streptococcus mitis*]; Sequence 4 (S4): SEQ ID NO: 29 | 5AXW_A | gi 924443546 | *Staphylococcus aureus* Cas9. The HNH domain (bold

and underlined) and the RuvC domain (boxed) are identified for each of the four sequences. Amino acid residues 10 and 840 in S1 and the homologous amino acids in the aligned sequences are identified with an asterisk following the respective amino acid residue.

```
S1 1    --MDKK-YSIGLD*IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLI--GALLFDSG--ETAEATRLKRTARRRYT  73
S2 1    --MTKKNYSIGLD*IGTNSVGWAVITDDYKVPAKKMKVLGNTDKKYIKKNLL--GALLFDSG--ETAEATRLKRTARRRYT  74
S3 1    --M-KKGYSIGLD*IGTNSVGFAVITDDYKVPSKKMKVLGNTDKRFIKKNLI--GALLFDEG--TTAEARRLKRTARRRYT  73
S4 1    GSHMKRNYILGLD*IGITSVGYGII--DYET-----------------RDVIDAGVRLFKEANVENNEGRRSKRGARRLKR  61

S1 74   RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRL 153
S2 75   RRKNRLRYLQEIFANEIAKVDESFFQRLDESFLTDDDKTFDSHPIFGNKAEEDAYHQKFPTIYHLRKHLADSSEKADLRL 154
S3 74   RRKNRLRYLQEIFSEEMSKVDSSFFHRLDDSFLIPEDKRESKYPIFATLTEEKEYHKQFPTIYHLRKQLADSKEKTDLRL 153
S4 62   RRRHRIQRVKKLL---------------FDYNLLTD---------------------HSELSGINPYEARVKGLSQKLSEEE 107

S1 154  IYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEK  233
S2 155  VYLALAHMIKFRGHFLIEGELNAENTDVQKIFADFVGVYNRTFDDSHLSEITVDVASILTEKISKSRRLENLIKYYPTEK  234
S3 154  IYLALAHMIKYRGHFLYEEAFDIKNNDIQKIFNEFISIYDNTFEGSSLSGQNAQVEAIFTDKISKSAKRERVLKLFPDEK  233
S4 108  FSAALLHLAKRRG----------------------VHNVNEVEEDT----------------------------------  131

S1 234  KNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEIT 313
S2 235  KNTLFGNLIALALGLQPNFKTNFKLSEDAKLQFSKDTYEEDLEELLGKIGDDYADLFTSAKNLYDAILLSGILTVDDNST 314
S3 234  STGLFSEFLKLIVGNQADFKKHFDLEDKAPLQFSKDTYDEDLENLLGQIGDDFTDLFVSAKKLYDAILLSGILTVTDPST 313
S4 132  -----GNELS-----------------TKEQISRN---------------------------------------------- 144

S1 314  KAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKM--DGTEELLV  391
S2 315  KAPLSASMIKRYVEHHEDLEKLKEFIKANKSELYHDIFKDKNKNGYAGYIENGVKQDEFYKYLKNILSKIKIDGSDYFLD  394
S3 314  KAPLSASMIERYENHQNDLAALKQFIKNNLPEKYDEVFSDQSKDGYAGYIDGKTTQETFYKYIKNLLSKF--EGTDYFLD  391
S4 145  ----SKALEEKYVAELQ-----------------------------------------------LERLKKDG------  165

S1 392  KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEE  471
S2 395  KIEREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGDYYPFLKEKQDRIEKILTFRIPYYVGPLVRKDSRFAWAEYRSDE  474
S3 392  KIEREDFLRKQRTFDNGSIPHQIHLQEMNAILRRQGEYYPFLKDNKEKIEKILTFRIPYYVGPLARGNRDFAWLTRNSDE  471
S4 166  --EVRGSINRFKTSD---------YVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGP--GEGSPFGW------K  227

S1 472  TITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL 551
S2 475  KITPWNFDKVIDKEKSAEKFITRMTLNDLYLPEEKVLPKHSHVYETYAVYNELTKIKYVNEQGKE-SFFDSNMKQEIFDH 553
S3 472  AIRPWNFEEIVDKASSAEDFINKMTNYDLYLPEEKVLPKHSLLYETFAVYNELTKVKFIAEGLRDYQFLDSGQKKQIVNQ 551
S4 228  DIKEW----------------YEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEK---LEYYEKFQIIEN 289

S1 552  LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR---FNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFED  628
S2 554  VFKENRKVTKEKLLNYLNKEFPEYRIKDLIGLDKENKSFNASLGTYHDLKKIL-DKAFLDDKVNEEVIEDIIKTLTLFED  632
S3 552  LFKENRKVTEKDIIHYLHN-VDGYDGIELKGIEKQ---FNASLSTYHDLLKIIKDKEFMDDAKNEAILENIVHTLTIFED  627
S4 290  VFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEF---TNLKVYHDIKDITARKEII---ENAELLDQIAKILTIYQS  363

S1 629  REMIEERLKTYAHLFDDKVMKQLKR-RRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKED 707
S2 633  KDMIHERLQKYSDIFTANQLKKLER-RHYTGWGRLSYKLINGIRNKENNKTILDYLIDDGSANRNFMQLINDDTLPFKQI 711
S3 628  REMIKQRLAQYDSLFDEKVIKALTR-RHYTGWGKLSAKLINGICDKQTGNTILDYLIDDGKINRNFMQLINDDGLSFKEI 706
S4 364  SEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDE------LWHTNDNQIAIFNRLKLVP---------- 428
```

```
S1 708  IQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT------QKGQKNSRERM 781
S2 712  IQKSQVVGDVDDIEAVVHDLPGSPAIKKGILQSVKIVDELVKVMG-GNPDNIVIEMARENQTT------NRGRSQSQQRL 784
S3 707  IQKAQVIGKTDDVKQVVQELSGSPAIKKGILQSIKIVDELVKVMG-HAPESIVIEMARENQTT------ARGKKNSQQRY 779
S4 429  -KKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYG--LPNDIIIELAREKNSKDAQKMINEMQKRNRQTN 505

S1 782  KRIEEGIKELGSQIL-------KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSD----YDVDH*IVPQSFLKDD 850
S2 785  KKLQNSLKELGSNILNEEKPSYIEDKVENSHLQNDQLFLYYIQNGKDMYTGDELDIDHLSD----YDIDH*IIPQAFIKDD 860
S3 780  KRIEDSLKILASGL---DSNILKENPTDNNQLQNDRLFLYYLQNGKDMYTGEALDINQLSS----YDIDH*IIPQAFIKDD 852
S4 506  ERIEEIIRTTGK---------------ENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDH*IIPRSVSFDN 570

S1 851  SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDN-LTKAERGGL-SELD------KAGFIKRQLV 922
S2 861  SIDNRVLTSSAKNRGKSDDVPSLDIVRARKAEWVRLYKSGLISKRKFDN-LTKAERGGL-TEAD------KAGFIKRQLV 932
S3 853  SLDNRVLTSSKDNRGKSDNVPSIEVVQKRKAFWQQLLDSKLISERKFNN-LTKAERGGL-DERD------KVGFIKRQLV 924
S4 571  SFNNKVLVKQEEASKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLV 650

S1 923  ETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYP 1002
S2 933  ETRQITKHVAQILDARFNTEHDENDKVIRDVKVITLKSNLVSQFRKDFEFYKVREINDYHHAHDAYLNAVVGTALLKKYP 1012
S3 925  ETRQITKHVAQILDARYNTEVNEKDKKNRTVKIITLKSNLVSNFRKEFRLYKVREINDYHHAHDAYLNAVVAKAILKKYP 1004
S4 651  DTRYATRGLMNLLRSYFRVN-------NLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIA---------- 712

S1 1003 KLESEFVYGDYKVYDVRKMIAKSEQ--EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG--- 1077
S2 1013 KLASEFVYGEYKKYDIRKFITNSSD-----KATAKYFFYSNLMNFFKTKVKYADGTVFERPIIETNAD-GEIAWNKQ--- 1083
S3 1005 KLEPEFVYGEYQKYDLKRYISRSKDPKEVEKATEKYFFYSNLLNFFKEEVHYADGTIVKRENIEYSKDTGEIAWNKE--- 1081
S4 713  --NADFIFKEWKKLDKAKKVMENQM-------------------------FEEKQAESMPEIETEQEYKEIFITPHQIK 764

S1 1078 -----RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD---WDPKKYGGFDSPTVAYSVLVVAKV 1149
S2 1084 -----IDFEKVRKVLSYPQVNIVKKVETQTGGFSKESILPKGDSDKLIPRKTKKVYWDTKKYGGFDSPTVAYSVFVVADV 1158
S3 1082 -----KDFAIIKKVLSLPQVNIVKKREVQTGGFSKESILPKGNSDKLIPRKTKDILLDTTKYGGFDSPVIAYSILLIADI 1156
S4 765  HIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKL----KKLIN-KSP----EKLLMYHH 835

S1 1150 EKGKSKKLKSVKELLGITIMERSSFEKNPI-DFLEAKG-----YKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKG 1223
S2 1159 EKGKAKKLKTVKELVGISIMERSFFEENPV-EFLENKG-----YHNIREDKLIKLPKYSLFEFEGGRRRLLASASELQKG 1232
S3 1157 EKGKAKKLKTVKTLVGITIMEKAAFEENPI-TFLENKG-----YHNVRKENILCLPKYSLFELENGRRRLLASAKELQKG 1230
S4 836  DPQTYQKLK--------LIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKV 907

S1 1224 NELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH------ 1297
S2 1233 NEMVLPGYLVELLYHAHRADNF-----NSTEYLNYVSEHKKEFEKVLSCVEDFANLYVDVEKNLSKIRAVADSM------ 1301
S3 1231 NEIVLPVYLTTLLYHSKNVHKL-----DEPGHLEYIQKHRNEFKDLLNLVSEFSQKYVLADANLEKIKSLYADN------ 1299
S4 908  VKLSLKPYRFD-VYLDNGVYKFV-----TVKNLDVIK--KENYYEVNSKAYEEAKKLKKISNQAEFIASFYNNDLIKING 979

S1 1298 RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSIT--------GLYETRI----DLSQL 1365
S2 1302 DNFSIEEISNSFINLLTLTALGAPADFNFLGEKIPRKRYTSTKECLNATLIHQSIT--------GLYETRI----DLSKL 1369
S3 1300 EQADIEILANSFINLLTFTALGAPAAFKFFGKDIDRKRYTTVSEILNATLIHQSIT--------GLYETWI----DLSKL 1367
S4 980  ELYRVIGVNNDLLNRIEVNMIDITYR-EYLENMNDKRPPRIIKTIASKT---QSIKKYSTDILGNLYEVKSKKHPQIIKK 1055
```

```
S1 1366 GGD  1368
S2 1370 GEE  1372
S3 1368 GED  1370
S4 1056 G--  1056
```

[00153] The alignment demonstrates that amino acid sequences and amino acid residues that are homologous to a reference Cas9 amino acid sequence or amino acid residue can be identified across Cas9 sequence variants, including, but not limited to Cas9 sequences from different species, by identifying the amino acid sequence or residue that aligns with the reference sequence or the reference residue using alignment programs and algorithms known in the art. This disclosure provides Cas9 variants in which one or more of the amino acid residues identified by an asterisk in SEQ ID NOs: 1 and 27-29 (*e.g.,* S1, S2, S3, and S4, respectively) are mutated as described herein. The residues D10 and H840 in Cas9 of SEQ ID NO: 1 that correspond to the residues identified in SEQ ID NOs: 1 and 27-29 by an asterisk are referred to herein as "homologous" or "corresponding" residues. Such homologous residues can be identified by sequence alignment, *e.g.*, as described above, and by identifying the sequence or residue that aligns with the reference sequence or residue. Similarly, mutations in Cas9 sequences that correspond to mutations identified in SEQ ID NO: 1 herein, *e.g.*, mutations of residues 10, and 840 in SEQ ID NO: 1, are referred to herein as "homologous" or "corresponding" mutations. For example, the mutations corresponding to the D10A mutation in SEQ ID NO: 1 (S1) for the four aligned sequences above are D11A for S2, D10A for S3, and D13A for S4; the corresponding mutations for H840A in SEQ ID NO: 1 (S1) are H850A for S2, H842A for S3, and H560A for S4.

[00154] A total of 250 Cas9 sequences (SEQ ID NOs: 1 and 27-275) from different species are provided. Amino acid residues corresponding to residues 10 and 840 of SEQ ID NO: 1 may be identified in the same manner as outlined above. All of these Cas9 sequences may be used in accordance with the present disclosure.

WP_010922251.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus pyogenes] SEQ ID NO: 1

WP_039695303.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus gallolyticus] SEQ ID NO: 27

WP_045635197.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mitis] SEQ ID NO: 28

5AXW_A Cas9, Chain A, Crystal Structure [Staphylococcus Aureus] SEQ ID NO: 29

WP_009880683.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus pyogenes] SEQ ID NO: 30

WP_010922251.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus pyogenes] SEQ ID NO: 31

WP_011054416.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus pyogenes] SEQ ID NO: 32

WP_011284745.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus pyogenes] SEQ ID NO: 33

WP_011285506.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus pyogenes] SEQ ID NO: 34

WP_011527619.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus pyogenes] SEQ ID NO: 35

WP_012560673.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus pyogenes] SEQ ID NO: 36

WP_014407541.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus pyogenes] SEQ ID NO: 37

WP_020905136.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus pyogenes] SEQ ID NO: 38

WP_023080005.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus pyogenes] SEQ ID NO: 39

WP_023610282.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus pyogenes] SEQ ID NO: 40

WP_030125963.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus pyogenes] SEQ ID NO: 41

WP_030126706.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus pyogenes] SEQ ID NO: 42

WP_031488318.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus pyogenes] SEQ ID NO: 43

WP_032460140.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus pyogenes] SEQ ID NO: 44

WP_032461047.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus pyogenes] SEQ ID NO: 45

WP_032462016.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus pyogenes] SEQ ID NO: 46

WP_032462936.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus pyogenes] SEQ ID NO: 47

WP_032464890.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus pyogenes] SEQ ID NO: 48

WP_033888930.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus pyogenes] SEQ ID NO: 49

WP_038431314.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus pyogenes] SEQ ID NO: 50

WP_038432938.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus pyogenes] SEQ ID NO: 51

WP_038434062.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus pyogenes] SEQ ID NO: 52

EAQ51233.1 CRISPR-associated protein, Csn1 family [Streptococcus pyogenes] SEQ ID NO: 53

KGE60162.1 hypothetical protein MGAS2111_0903 [Streptococcus pyogenes MGAS2111] SEQ ID NO: 54

KGE60856.1 CRISPR-associated endonuclease protein [Streptococcus pyogenes SS1447] SEQ ID NO: 55

WP_002989955.1 MULTISPECIES: type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus] SEQ ID NO: 56

WP_003030002.1 MULTISPECIES: type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus] SEQ ID NO: 57

WP_003065552.1 MULTISPECIES: type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus] SEQ ID NO: 58

WP_001040076.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 59

WP_001040078.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 60

WP_001040080.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 61

WP_001040081.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 62

WP_001040083.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 63

WP_001040085.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 64

WP_001040087.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 65

WP_001040088.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 66

WP_001040089.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 67

WP_001040090.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 68

WP_001040091.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 69

WP_001040092.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 70

WP_001040094.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 71

WP_001040095.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 72

WP_001040096.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 73

WP_001040097.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 74

WP_001040098.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 75

WP_001040099.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 76

WP_001040100.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 77

WP_001040104.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 78

WP_001040105.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 79

WP_001040106.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 80

WP_001040107.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 81

WP_001040108.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 82

WP_001040109.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 83

WP_001040110.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 84

WP_015058523.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 85

WP_017643650.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 86

WP_017647151.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 87

WP_017648376.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 88

WP_017649527.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 89

WP_017771611.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 90

WP_017771984.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 91

CFQ25032.1 CRISPR-associated protein [Streptococcus agalactiae] SEQ ID NO: 92

CFV16040.1 CRISPR-associated protein [Streptococcus agalactiae] SEQ ID NO: 93

KLJ37842.1 CRISPR-associated protein Csn1 [Streptococcus agalactiae] SEQ ID NO: 94

KLJ72361.1 CRISPR-associated protein Csn1 [Streptococcus agalactiae] SEQ ID NO: 95

KLL20707.1 CRISPR-associated protein Csn1 [Streptococcus agalactiae] SEQ ID NO: 96

KLL42645.1 CRISPR-associated protein Csn1 [Streptococcus agalactiae] SEQ ID NO: 97

WP_047207273.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 98

WP_047209694.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 99

WP_050198062.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 100

WP_050201642.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 101

WP_050204027.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 102

WP_050881965.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 103

WP_050886065.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus agalactiae] SEQ ID NO: 104

AHN30376.1 CRISPR-associated protein Csn1 [Streptococcus agalactiae 138P] SEQ ID NO: 105

EAO78426.1 reticulocyte binding protein [Streptococcus agalactiae H36B] SEQ ID NO: 106

CCW42055.1 CRISPR-associated protein, SAG0894 family [Streptococcus agalactiae ILRI112] SEQ ID NO:107

WP_003041502.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus anginosus] SEQ ID NO: 108

WP_037593752.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus anginosus] SEQ ID NO: 109

WP_049516684.1 CRISPR-associated protein Csn1 [Streptococcus anginosus] SEQ ID NO: 110

GAD46167.1 hypothetical protein ANG6_0662 [Streptococcus anginosus T5] SEQ ID NO: 111

WP_018363470.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus caballi] SEQ ID NO: 112

WP_003043819.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus canis] SEQ ID NO: 113

WP_006269658.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus constellatus] SEQ ID NO: 114

WP_048800889.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus constellatus] SEQ ID NO: 115

WP_012767106.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus dysgalactiae] SEQ ID NO: 116

WP_014612333.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus dysgalactiae] SEQ ID NO: 117

WP_015017095.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus dysgalactiae] SEQ ID NO: 118

WP_015057649.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus dysgalactiae] SEQ ID NO: 119

WP_048327215.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus dysgalactiae] SEQ ID NO: 120

WP_049519324.1 CRISPR-associated protein Csn1 [Streptococcus dysgalactiae] SEQ ID NO: 121

WP_012515931.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus equi] SEQ ID NO: 122

WP_021320964.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus equi] SEQ ID NO: 123

WP_037581760.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus equi] SEQ ID NO: 124

WP_004232481.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus equinus] SEQ ID NO: 125

WP_009854540.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus gallolyticus] SEQ ID NO: 126

WP_012962174.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus gallolyticus] SEQ ID NO: 127

WP_039695303.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus gallolyticus] SEQ ID NO: 128

WP_014334983.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus infantarius] SEQ ID NO: 129

WP_003099269.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus iniae] SEQ ID NO: 130

AHY15608.1 CRISPR-associated protein Csn1 [Streptococcus iniae] SEQ ID NO: 131

AHY17476.1 CRISPR-associated protein Csn1 [Streptococcus iniae] SEQ ID NO: 132

ESR09100.1 hypothetical protein IUSA1_08595 [Streptococcus iniae IUSA1] SEQ ID NO: 133

AGM98575.1 CRISPR-associated protein Cas9/Csn1, subtype II/NMEMI [Streptococcus iniae SF1] SEQ ID NO: 134

ALF27331.1 CRISPR-associated protein Csn1 [Streptococcus intermedius] SEQ ID NO: 135

WP_018372492.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus massiliensis] SEQ ID NO: 136

WP_045618028.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mitis] SEQ ID NO: 137

WP_045635197.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mitis] SEQ ID NO: 138

WP_002263549.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 139

WP_002263887.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 140

WP_002264920.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 141

WP_002269043.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 142

WP_002269448.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 143

WP_002271977.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 144

WP_002272766.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 145

WP_002273241.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 146

WP_002275430.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 147

WP_002276448.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 148

WP_002277050.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 149

WP_002277364.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 150

WP_002279025.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 151

WP_002279859.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 152

WP_002280230.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 153

WP_002281696.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 154

WP_002282247.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 155

WP_002282906.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 156

WP_002283846.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 157

WP_002287255.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 158

WP_002288990.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 159

WP_002289641.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 160

WP_002290427.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 161

WP_002295753.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 162

WP_002296423.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 163

WP_002304487.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 164

WP_002305844.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 165

WP_002307203.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 166

WP_002310390.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 167

WP_002352408.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 168

WP_012997688.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 169

WP_014677909.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 170

WP_019312892.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 171

WP_019313659.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 172

WP_019314093.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 173

WP_019315370.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 174

WP_019803776.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 175

WP_019805234.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 176

WP_024783594.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 177

WP_024784288.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 178

WP_024784666.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 179

WP_024784894.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 180

WP_024786433.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mutans] SEQ ID NO: 181

WP_049473442.1 CRISPR-associated protein Csn1 [Streptococcus mutans] SEQ ID NO: 182

WP_049474547.1 CRISPR-associated protein Csn1 [Streptococcus mutans] SEQ ID NO: 183

EMC03581.1 hypothetical protein SMU69_09359 [Streptococcus mutans NLML4] SEQ ID NO: 184

WP_000428612.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus oralis] SEQ ID NO: 185

WP_000428613.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus oralis] SEQ ID NO: 186

WP_049523028.1 CRISPR-associated protein Csn1 [Streptococcus parasanguinis] SEQ ID NO: 187

WP_003107102.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus parauberis] SEQ ID NO: 188

WP_054279288.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus phocae] SEQ ID NO: 189

WP_049531101.1 CRISPR-associated protein Csn1 [Streptococcus pseudopneumoniae] SEQ ID NO: 190

WP_049538452.1 CRISPR-associated protein Csn1 [Streptococcus pseudopneumoniae] SEQ ID NO: 191

WP_049549711.1 CRISPR-associated protein Csn1 [Streptococcus pseudopneumoniae] SEQ ID NO: 192

WP_007896501.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus pseudoporcinus] SEQ ID NO: 193

EFR44625.1 CRISPR-associated protein, Csn1 family [Streptococcus pseudoporcinus SPIN 20026] SEQ ID NO: 194

WP_002897477.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus sanguinis] SEQ ID NO: 195

WP_002906454.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus sanguinis] SEQ ID NO: 196

WP_009729476.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus sp. F0441] SEQ ID NO: 197

CQR24647.1 CRISPR-associated protein [Streptococcus sp. FF10] SEQ ID NO: 198

WP_000066813.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus sp. M334] SEQ ID NO: 199

WP_009754323.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus sp. taxon 056] SEQ ID NO: 200

WP_044674937.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus suis] SEQ ID NO: 201

WP_044676715.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus suis] SEQ ID NO: 202

WP_044680361.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus suis] SEQ ID NO: 203

WP_044681799.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus suis] SEQ ID NO: 204

WP_049533112.1 CRISPR-associated protein Csn1 [Streptococcus suis] SEQ ID NO: 205

WP_029090905.1 type II CRISPR RNA-guided endonuclease Cas9 [Brochothrix thermosphacta] SEQ ID NO: 206

WP_006506696.1 type II CRISPR RNA-guided endonuclease Cas9 [Catenibacterium mitsuokai] SEQ ID NO: 207

AIT42264.1 Cas9hc:NLS:HA [Cloning vector pYB196] SEQ ID NO: 208

WP_034440723.1 type II CRISPR endonuclease Cas9 [Clostridiales bacterium S5-A11] SEQ ID NO: 209

AKQ21048.1 Cas9 [CRISPR-mediated gene targeting vector p(bhsp68-Cas9)] SEQ ID NO: 210

WP_004636532.1 type II CRISPR RNA-guided endonuclease Cas9 [Dolosigranulum pigrum] SEQ ID NO: 211

WP_002364836.1 MULTISPECIES: type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus] SEQ ID NO: 212

WP_016631044.1 MULTISPECIES: type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus] SEQ ID NO: 213

EMS75795.1 hypothetical protein H318_06676 [Enterococcus durans IPLA 655] SEQ ID NO: 214

WP_002373311.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus faecalis] SEQ ID NO: 215

WP_002378009.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus faecalis] SEQ ID NO: 216

WP_002407324.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus faecalis] SEQ ID NO: 217

WP_002413717.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus faecalis] SEQ ID NO: 218

WP_010775580.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus faecalis] SEQ ID NO: 219

WP_010818269.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus faecalis] SEQ ID NO: 220

WP_010824395.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus faecalis] SEQ ID NO: 221

WP_016622645.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus faecalis] SEQ ID NO: 222

WP_033624816.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus faecalis] SEQ ID NO: 223

WP_033625576.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus faecalis] SEQ ID NO: 224

WP_033789179.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus faecalis] SEQ ID NO: 225

WP_002310644.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus faecium] SEQ ID NO: 226

WP_002312694.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus faecium] SEQ ID NO: 227

WP_002314015.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus faecium] SEQ ID NO: 228

WP_002320716.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus faecium] SEQ ID NO: 229

WP_002330729.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus faecium] SEQ ID NO: 230

WP_002335161.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus faecium] SEQ ID NO: 231

WP_002345439.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus faecium] SEQ ID NO: 232

WP_034867970.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus faecium] SEQ ID NO: 233

WP_047937432.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus faecium] SEQ ID NO: 234

WP_010720994.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus hirae] SEQ ID NO: 235

WP_010737004.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus hirae] SEQ ID NO: 236

WP_034700478.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus hirae] SEQ ID NO: 237

WP_007209003.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus italicus] SEQ ID NO: 238

WP_023519017.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus mundtii] SEQ ID NO: 239

WP_010770040.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus phoeniculicola] SEQ ID NO: 240

WP_048604708.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus sp. AM1] SEQ ID NO: 241

WP_010750235.1 type II CRISPR RNA-guided endonuclease Cas9 [Enterococcus villorum] SEQ ID NO: 242

AII16583.1 Cas9 endonuclease [Expression vector pCas9] SEQ ID NO: 243

WP_029073316.1 type II CRISPR RNA-guided endonuclease Cas9 [Kandleria vitulina] SEQ ID NO: 244

WP_031589969.1 type II CRISPR RNA-guided endonuclease Cas9 [Kandleria vitulina] SEQ ID NO: 245

KDA45870.1 CRISPR-associated protein Cas9/Csn1, subtype II/NMEMI [Lactobacillus animalis] SEQ ID NO: 246

WP_039099354.1 type II CRISPR RNA-guided endonuclease Cas9 [Lactobacillus curvatus] SEQ ID NO: 247

AKP02966.1 hypothetical protein ABB45_04605 [Lactobacillus farciminis] SEQ ID NO: 248

WP_010991369.1 type II CRISPR RNA-guided endonuclease Cas9 [Listeria innocua] SEQ ID NO: 249

WP_033838504.1 type II CRISPR RNA-guided endonuclease Cas9 [Listeria innocua] SEQ ID NO: 250

EHN60060.1 CRISPR-associated protein, Csn1 family [Listeria innocua ATCC 33091] SEQ ID NO: 251

EFR89594.1 crispr-associated protein, Csn1 family [Listeria innocua FSL S4-378] SEQ ID NO: 252

WP_038409211.1 type II CRISPR RNA-guided endonuclease Cas9 [Listeria ivanovii] SEQ ID NO: 253

EFR95520.1 crispr-associated protein Csn1 [Listeria ivanovii FSL F6-596] SEQ ID NO: 254

WP_003723650.1 type II CRISPR RNA-guided endonuclease Cas9 [Listeria monocytogenes] SEQ ID NO: 255

WP_003727705.1 type II CRISPR RNA-guided endonuclease Cas9 [Listeria monocytogenes] SEQ ID NO: 256

WP_003730785.1 type II CRISPR RNA-guided endonuclease Cas9 [Listeria monocytogenes] SEQ ID NO: 257

WP_003733029.1 type II CRISPR RNA-guided endonuclease Cas9 [Listeria monocytogenes] SEQ ID NO: 258

WP_003739838.1 type II CRISPR RNA-guided endonuclease Cas9 [Listeria monocytogenes] SEQ ID NO: 259

WP_014601172.1 type II CRISPR RNA-guided endonuclease Cas9 [Listeria monocytogenes] SEQ ID NO: 260

WP_023548323.1 type II CRISPR RNA-guided endonuclease Cas9 [Listeria monocytogenes] SEQ ID NO: 261

WP_031665337.1 type II CRISPR RNA-guided endonuclease Cas9 [Listeria monocytogenes] SEQ ID NO: 262

WP_031669209.1 type II CRISPR RNA-guided endonuclease Cas9 [Listeria monocytogenes] SEQ ID NO: 263

WP_033920898.1 type II CRISPR RNA-guided endonuclease Cas9 [Listeria monocytogenes] SEQ ID NO: 264

AKI42028.1 CRISPR-associated protein [Listeria monocytogenes] SEQ ID NO: 265

AKI50529.1 CRISPR-associated protein [Listeria monocytogenes] SEQ ID NO: 266

EFR83390.1 crispr-associated protein Csn1 [Listeria monocytogenes FSL F2-208] SEQ ID NO: 267

WP_046323366.1 type II CRISPR RNA-guided endonuclease Cas9 [Listeria seeligeri] SEQ ID NO: 268

AKE81011.1 Cas9 [Plant multiplex genome editing vector pYLCRISPR/Cas9Pubi-H] SEQ ID NO: 269

CUO82355.1 Uncharacterized protein conserved in bacteria [Roseburia hominis] SEQ ID NO: 270

WP_033162887.1 type II CRISPR RNA-guided endonuclease Cas9 [Sharpea azabuensis] SEQ ID NO: 271

AGZ01981.1 Cas9 endonuclease [synthetic construct] SEQ ID NO: 272

AKA60242.1 nuclease deficient Cas9 [synthetic construct] SEQ ID NO: 273

AKS40380.1 Cas9 [Synthetic plasmid pFC330] SEQ ID NO: 274

4UN5_B Cas9, Chain B, Crystal Structure SEQ ID NO: 275

[00155] Non-limiting examples of suitable cytosine deaminase domains are provided.

Human AID

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDWD
LDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMT
FKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 276)

Mouse AID

MDSLLMKQKKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSCSLDFGHLRNKSGCHVELLFLRYISDWD
LDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIGIMT
FKDYFYCWNTFVENRERTFKAWEGLHENSVRLTRQLRRILLPLYEVDDLRDAFRMLGF (SEQ ID NO: 277)

Dog AID

MDSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGCHVELLFLRYISDWD
LDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAARLYFCEDRKAEPEGLRRLHRAGVQIAIMT
FKDYFYCWNTFVENREKTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 278)

Bovine AID

MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGCHVELLFLRYISDWD
LDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFCDKERKAEPEGLRRLHRAGVQIAIM
TFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO:
279)

Mouse APOBEC-3

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLHHGVFKNKDNIH
AEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHHNLSLDIFSSRLYNVQDPETQQNL
CRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKRLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSN
ICLTKGLPETRFCVEGRRMDPLSEEEFYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEK
GKQHAEILFLDKIRSMELSQVTITCYLTWSPCPNCAWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCS
LWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLRRIKESWGLQDLVNDFGNLQLGP
PMS (SEQ ID NO: 280)

Rat APOBEC-3

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLRYAIDRKDTFLCYEVTRKDCDSPVSLHHGVFKNKDNIHA
EICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQVLRFLATHHNLSLDIFSSRLYNIRDPENQQNLC
RLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKKLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNI
CLTKGLPETRFCVERRRVHLLSEEEFYSQFYNQRVKHLCYYHGVKPYLCYQLEQFNGQAPLKGCLLSEKG
KQHAEILFLDKIRSMELSQVIITCYLTWSPCPNCAWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSL
WQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLHRIKESWGLQDLVNDFGNLQLGPP
MS (SEQ ID NO: 281)

Rhesus macaque APOBEC-3G

MVEPMDPRTFVSNFNNRPILSGLNTVWLCCEVKTKDPSGPPLDAKIFQGKVYSKAKYHPEMRFLRWFHKW
RQLHHDQEYKVTWYVSWSPCTRCANSVATFLAKDPKVTLTIFVARLYYFWKPDYQQALRILCQKRGGPH

ATMKIMNYNEFQDCWNKFVDGRGKPFKPRNNLPKHYTLLQATLGELLRHLMDPGTFTSNFNNKPWVSGQ
HETYLCYKVERLHNDTWVPLNQHRGFLRNQAPNIHGFPKGRHAELCFLDLIPFWKLDGQQYRVTCFTSWS
PCFSCAQEMAKFISNNEHVSLCIFAARIYDDQGRYQEGLRALHRDGAKIAMMNYSEFEYCWDTFVDRQGR
PFQPWDGLDEHSQALSGRLRAI (SEQ ID NO: 282) (italic: nucleic acid editing domain; underline: cytoplasmic localization signal)

Chimpanzee APOBEC-3G

MKPHFRNPVERMYQDTFSDNFYNRPILSHRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSKLKYHPEMRF
FHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDVATFLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQ
KRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTSNFNNELW
VRGRHETYLCYEVERLHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLHQDYRVT
CFTSWSPCFSCAQEMAKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLAKAGAKISIMTYSEFKHCWDTFV
DHQGCPFQPWDGLEEHSQALSGRLRAILQNQGN (SEQ ID NO: 283)

Green monkey APOBEC-3G

MNPQIRNMVEQMEPDIFVYYFNNRPILSGRNTVWLCYEVKTKDPSGPPLDANIFQGKLYPEAKDHPEMKFL
HWFRKWRQLHRDQEYEVTWYVSWSPCTRCANSVATFLAEDPKVTLTIFVARLYYFWKPDYQQALRILCQ
ERGGPHATMKIMNYNEFQHCWNEFVDGQGKPFKPRKNLPKHYTLLHATLGELLRHVMDPGTFTSNFNNK
PWVSGQRETYLCYKVERSHNDTWVLLNQHRGFLRNQAPDRHGFPKGRHAELCFLDLIPFWKLDDQQYRV
TCFTSWSPCFSCAQKMAKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLHRDGAKIAVMNYSEFEYCWDTF
VDRQGRPFQPWDGLDEHSQALSGRLRAI (SEQ ID NO: 284)

Human APOBEC-3G

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSELKYHPEMRFF
HWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQ
KRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPW
VRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVT
CFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVD
HQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN (SEQ ID NO: 285)

Human APOBEC-3F

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQVYSQPEHHAEMCF
LSWFCGNQLPAYKCFQITWFVSWTPCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERDYRRALCRLSQA
GARVKIMDDEEFAYCWENFVYSEGQPFMPWYKFDDNYAFLHRTLKEILRNPMEAMYPHIFYFHFKNLRK
AYGRNESWLCFTMEVVKHHSPVSWKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWS
PCPECAGEVAEFLARHSNVNLTIFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCWENFVYNDD
EPFKPWKGLKYNFLFLDSKLQEILE (SEQ ID NO: 286)

Human APOBEC-3B

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGQVYFKPQYHAEM
CFLSWFCGNQLPAYKCFQITWFVSWTPCPDCVAKLAEFLSEHPNVTLTISAARLYYYWERDYRRALCRLSQ
AGARVTIMDYEEFAYCWENFVYNEGQQFMPWYKFDENYAFLHRTLKEILRYLMDPDTFTFNFNNDPLVL
RRRQTYLCYEVERLDNGTWVLMDQHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWF
ISWSPCFSWGCAGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTF
VYRQGCPFQPWDGLEEHSQALSGRLRAILQNQGN (SEQ ID NO: 287)

Human APOBEC-3C:

MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRNQVDSETHCHAER
CFLSWFCDDILSPNTKYQVTWYTSWSPCPDCAGEVAEFLARHSNVNLTIFTARLYYFQYPCYQEGLRSLSQ
EGVAVEIMDYEDFKYCWENFVYNDNEPFKPWKGLKTNFRLLKRRLRESLQ (SEQ ID NO: 288)

Human APOBEC-3A:

MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQAKNLLCGFYGRH
AELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQ
MLRDAGAQVSIMTYDEFKHCWDTFVDHQGCPFQPWDGLDEHSQALSGRLRAILQNQGN (SEQ ID NO:
289)

Human APOBEC-3H:

MALLTAETFRLQFNNKRRLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKCHAEICFINEIKSMGLDET
QCYQVTCYLTWSPCSSCAWELVDFIKAHDHLNLGIFASRLYYHWCKPQQKGLRLLCGSQVPVEVMGFPKF
ADCWENFVDHEKPLSFNPYKMLEELDKNSRAIKRRLERIKIPGVRAQGRYMDILCDAEV (SEQ ID NO:
290)

Human APOBEC-3D

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGPVLPKRQSNHRQE
VYFRFENHAEMCFLSWFCGNRLPANRRFQITWFVSWNPCLPCVVKVTKFLAEHPNVTLTISAARLYYYRD
RDWRWVLLRLHKAGARVKIMDYEDFAYCWENFVCNEGQPFMPWYKFDDNYASLHRTLKEILRNPMEAM
YPHIFYFHFKNLLKACGRNESWLCFTMEVTKHHSAVFRKRGVFRNQVDPETHCHAERCFLSWFCDDILSPN
TNYEVTWYTSWSPCPECAGEVAEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGASVKIMGYKDF
VSCWKNFVYSDDEPFKPWKGLQTNFRLLKRRLREILQ (SEQ ID NO: 291)

Human APOBEC-1

MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKKFTS
ERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVARLFWHMDQQNRQGLRDLVNSGVTIQI
MRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNC
HYQTIPPHILLATGLIHPSVAWR (SEQ ID NO: 292)

Mouse APOBEC-1

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSVWRHTSQNTSNHVEVNFLEKFTT
ERYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPYVTLFIYIARLYHHTDQRNRQGLRDLISSGVTIQIMTE
QEYCYCWRNFVNYPPSNEAYWPRYPHLWVKLYVLELYCIILGLPPCLKILRRKQPQLTFFTITLQTCHYQRI
PPHLLWATGLK (SEQ ID NO: 293)

Rat APOBEC-1

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTT
ERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTE
QESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRL
PPHILWATGLK (SEQ ID NO: 294)

Petromyzon marinus CDA1 (pmCDA1)

MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAEIFSI
RKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNL

RDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAV (SEQ ID NO: 295)

Human APOBEC3G D316R_D317R

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSELKYHPEMRFF HWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQ KRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPW VRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVT CFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYRRQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVD HQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN (SEQ ID NO: 296)

Human APOBEC3G chain A

MDPPTFTFNFNNEPWVVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDV IPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISI MTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQ (SEQ ID NO: 297)

Human APOBEC3G chain A D120R_D121R

MDPPTFTFNFNNEPWVVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDV IPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYRRQGRCQEGLRTLAEAGAKISI MTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQ (SEQ ID NO: 298)

[00156] Nonlimiting, exemplary uracil glycosylase inhibitor sequences are provided.

Bacillus phage PBS2 (Bacteriophage PBS2) Uracil-DNA glycosylase inhibitor
MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQ DSNGENKIKML (SEQ ID NO: 299)

Erwinia tasmaniensis SSB (themostable single-stranded DNA binding protein)

MASRGVNKVILVGNLGQDPEVRYMPNGGAVANITLATSESWRDKQTGETKEKTEWHRVVLFGKLAEVA GEYLRKGSQVYIEGALQTRKWTDQAGVEKYTTEVVVNVGGTMQMLGGRSQGGGASAGGQNGGSNNGW GQPQQPQGGNQFSGGAQQQARPQQQPQQNNAPANNEPPIDFDDDIP (SEQ ID NO: 300)

UdgX (binds to uracil in DNA but does not excise)

MAGAQDFVPHTADLAELAAAAGECRGCGLYRDATQAVFGAGGRSARIMMIGEQPGDKEDLAGLPFVGPA GRLLDRALEAADIDRDALYVTNAVKHFKFTRAAGGKRRIHKTPSRTEVVACRPWLIAEMTSVEPDVVVLL GATAAKALLGNDFRVTQHRGEVLHVDDVPGDPALVATVHPSSLLRGPKEERESAFAGLVDDLRVAADVR P (SEQ ID NO: 301)

UDG (catalytically inactive human UDG, binds to uracil in DNA but does not excise)

MIGQKTLYSFFSPSPARKRHAPSPEPAVQGTGVAGVPEESGDAAAIPAKKAPAGQEEPGTPPSSPLSAEQLD RIQRNKAAALLRLAARNVPVGFGESWKKHLSGEFGKPYFIKLMGFVAEERKHYTVYPPPHQVFTWTQMC DIKDVKVVILGQEPYHGPNQAHGLCFSVQRPVPPPPSLENIYKELSTDIEDFVHPGHGDLSGWAKQGVLLLN AVLTVRAHQANSHKERGWEQFTDAVVSWLNQNSNGLVFLLWGSYAQKKGSAIDRKRHHVLQTAHPSPLS VYRGFFGCRHFSKTNELLQKSGKKPIDWKEL (SEQ ID NO: 302)

[00157] Non-limiting examples of C to T nucleobase editors are provided.

His_6-rAPOBEC1-XTEN-dCas9 for *Escherichia coli* expression (SEQ ID NO: 303)

MGSSHHHHHHMSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKH
VEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLI
SSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTI
ALQSCHYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLG
NTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE
EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFK
SNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE
DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL
SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN
EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLK
SDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRH
KPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVD
QELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQR
KFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESI
LPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDF
LEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDN
EQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

rAPOBEC1-XTEN-dCas9-NLS for mammalian expression (SEQ ID NO: 304)

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTT
ERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTE
QESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRL
PPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN
LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQT
YNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKL
QLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK
ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN
GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEV
VDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL
LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT
LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFM
QLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARE
NQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDY
DVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG
GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREI
NNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKT
EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVK
KDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH
KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRY
TSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

hAPOBEC1-XTEN-dCas9-NLS for Mammalian expression (SEQ ID NO: 305)

MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKKFTS
ERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVARLFWHMDQQNRQGLRDLVNSGVTIQI
MRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNC
HYQTIPPHILLATGLIHPSVAWRSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVL
GNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLV
EEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNF
KSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKR
YDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNR
EDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRK
SEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF
LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE
NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFL
KSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGR
HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV
DQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQ
RKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDF
RKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAK
YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSK
ESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI
DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPE
DNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAA
FKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

rAPOBEC1-XTEN-dCas9-UGI-NLS (SEQ ID NO: 306)

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTT
ERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTE
QESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRL
PPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN
LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQT
YNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKL
QLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK
ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN
GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEV
VDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL
LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT
LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFM
QLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARE
NQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDY
DVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG
GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREI
NNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKT
EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVK
KDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH
KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRY
TSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESD
ILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV

rAPOBEC1-XTEN-Cas9 nickase-UGI-NLS (BE3, SEQ ID NO: 307)

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTT
ERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTE
QESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRL
PPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN
LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQT
YNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKL
QLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK
ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN
GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEV
VDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL
LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTITL
FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQ
LIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAREN
QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD
VDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN
NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTE
ITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIAR
KKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHK
HYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYT
STKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDI
LVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV

pmCDA1-XTEN-dCas9-UGI (bacteria) (SEQ ID NO: 308)

MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAEIFSI
RKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNL
RDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAVSGSET
PGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPT
IYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGV
DAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNL
LAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL
RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM
TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK
EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTY
AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQK
AQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSI
DNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL
VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD
SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN
GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKR
VILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI
TGLYETRIDLSQLGGDSGGSMTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDEN
VMLLTSDAPEYKPWALVIQDSNGENKIKML

pmCDA1-XTEN-nCas9-UGI-NLS (mammalian construct) (SEQ ID NO: 309)

MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAEIFSI
RKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNL
RDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAVSGSET
PGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPT
IYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGV
DAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNL
LAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL
RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM
TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK
EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTY
AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQK
AQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSI
DNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL
VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD
SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN
GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKR
VILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI
TGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENV
MLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV

huAPOBEC3G-XTEN-dCas9-UGI (bacteria) (SEQ ID NO: 310)

MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDV
IPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISI
MTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQSGSETPGTSESATPESDKKYSIGLAIGTN
SVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIF
SNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL
AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPG
EKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILL
SDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYK
FIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIP
YYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS
LGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGR
LSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIK
KGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQ
LQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEV
VKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE
NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV
YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKL
KSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPS
KYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSMTN
LSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSN
GENKIKML

huAPOBEC3G-XTEN-nCas9-UGI-NLS (mammalian construct) (SEQ ID NO: 311)

MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDV
IPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISI
MTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQSGSETPGTSESATPESDKKYSIGLAIGTN
SVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIF
SNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL
AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPG
EKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILL
SDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYK
FIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIP
YYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS
LGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGR
LSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIK
KGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQ
LQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEV
VKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE
NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV
YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKL
KSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPS
KYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLS
DIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGE
NKIKMLSGGSPKKKRKV

huAPOBEC3G (D316R_D317R)-XTEN-nCas9-UGI-NLS (mammalian construct) (SEQ ID NO: 312)

MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDV
IPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYRRQGRCQEGLRTLAEAGAKISI
MTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQSGSETPGTSESATPESDKKYSIGLAIGTN
SVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIF
SNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL
AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPG
EKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILL
SDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYK
FIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIP
YYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS
LGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGR
LSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIK
KGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQ
LQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEV
VKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE
NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV
YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKL
KSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPS
KYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLS
DIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGE
NKIKMLSGGSPKKKRKV

High fidelity nucleobase editor (SEQ ID NO: 313)

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTT
ERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTE
QESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRL
PPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN
LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQT
YNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKL
QLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK
ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN
GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEV
VDKGASAQSFIERMTAFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL
LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT
LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGALSRKLINGIRDKQSGKTILDFLKSDGFANRNFM
ALIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARE
NQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDY
DVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG
GLSELDKAGFIKRQLVETRAITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREI
NNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKT
EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVK
KDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH
KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRY
TSTKEVLDATLIHQSITGLYETRIDLSQLGGD

rAPOBEC1-XTEN-SaCas9n-UGI-NLS)

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTT
ERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTE
QESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRL
PPHILWATGLKSGSETPGTSESATPES<u>KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGR</u>
<u>RSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGV</u>
<u>HNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQK</u>
<u>AYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNA</u>
<u>LNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYH</u>
<u>DIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDE</u>
<u>LWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELARE</u>
<u>KNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNY</u>
<u>EVDHIIPRSVSFDNSFNNKVLVKQEEASKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYL</u>
<u>LEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNK</u>
<u>GYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKD</u>
<u>YKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQ</u>
<u>KLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLS</u>
<u>LKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELY</u>
<u>RVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG</u>
SGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWAL
VIQDSNGENKIKMLSGGSPKKKRKV (SEQ ID NO: 399)

rAPOBEC1-XTEN-SaCas9n-UGI-NLS

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTT
ERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTE
QESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRL
PPHILWATGLKSGSETPGTSESATPES<u>KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGR</u>
<u>RSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGV</u>
<u>HNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQK</u>

AYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNA
LNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYH
DIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDE
LWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELARE
KNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNY
EVDHIIPRSVSFDNSFNNKVLVKQEEASKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYL
LEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNK
GYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKD
YKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQ
KLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLS
LKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYKNDLIKINGELY
RVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG
SGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWAL
VIQDSNGENKIKMLSGGSPKKKRKV (SEQ ID NO: 400)

Nucleobase Editor 4-SSB

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTT
ERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTE
QESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRL
PPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN
LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQT
YNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKL
QLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK
ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN
GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEV
VDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL
LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT
LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFM
QLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARE
NQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDY
DVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG
GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREI
NNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKT
EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVK
KDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH
KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRY
TSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSGGSGGSASRGVNKVILVGNLGQDPEVRYMPNGGAV
ANITLATSESWRDKATGEMKEQTEWHRVVLFGKLAEVASEYLRKGSQVYIEGQLRTRKWTDQSGQDRYT
TEVVVNVGGTMQMLGGRQGGGAPAGGNIGGGQPQGGWGQPQQPQGGNQFSGGAQSRPQQSAPAAPSNE
PPMDFDDDIPFSGGSPKKKRKV (SEQ ID NO: 401)

Nucleobase Editor 4-$(GGS)_3$

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTT
ERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTE
QESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRL
PPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN
LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQT
YNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKL
QLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK
ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN
GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEV

VDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL
LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT
LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFM
QLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARE
NQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDY
DVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG
GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREI
NNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKT
EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVK
KDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH
KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRY
TSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIG
NKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV (SEQ ID NO:
402)

Nucleobase Editor 4-XTEN

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTT
ERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTE
QESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRL
PPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN
LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQT
YNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKL
QLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK
ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN
GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEV
VDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL
LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT
LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFM
QLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARE
NQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDY
DVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG
GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREI
NNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKT
EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVK
KDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH
KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRY
TSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGSETPGTSESATPESTNLSDIIEKETGKQLVIQESILMLPEE
VEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV (SEQ
ID NO: 403)

Nucleobase Editor 4-32aa linker

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTT
ERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTE
QESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRL
PPHILWATGLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKK
FKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEE
SFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLV

KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM
RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF
LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGK
TILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGR
DMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNA
KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFE
KNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLK
GSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLG
APAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESIL
MLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRK
V (SEQ ID NO: 404)

Nucleobase Editor 4-2X UGI

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTT
ERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTE
QESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRL
PPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN
LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQT
YNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKL
QLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK
ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN
GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEV
VDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL
LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT
LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFM
QLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARE
NQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDY
DVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG
GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREI
NNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKT
EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVK
KDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH
KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRY
TSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESD
ILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSTNLSDIIEKETGKQLVIQESILMLP
EEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV
(SEQ ID NO: 405)

Nucleobase Editor 4

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTT
ERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTE
QESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRL
PPHILWATGLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKK
FKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEE
SFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS

MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLV
KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM
RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF
LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGK
TILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGR
DMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNA
KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFE
KNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLK
GSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLG
APAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSGGSGGSTNLSDIIEKETGKQ
LVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGG
SGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKP
WALVIQDSNGENKIKMLSGGSPKKKRKV (SEQ ID NO: 406)

[00158] Non-limiting examples evolved adenosine deaminases that accept DNA as substrates are provided.

ecTadA
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILAD
ECAALLSDFFRMRRQEIKAQKKAQSSTD (SEQ ID NO: 314)

ecTadA (D108N)
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARNAKTGAAGSLMDVLHHPGMNHRVEITEGILAD
ECAALLSDFFRMRRQEIKAQKKAQSSTD (SEQ ID NO: 315)

ecTadA (D108G)
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARGAKTGAAGSLMDVLHHPGMNHRVEITEGILAD
ECAALLSDFFRMRRQEIKAQKKAQSSTD (SEQ ID NO: 316)

ecTadA (D108V)
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARVAKTGAAGSLMDVLHHPGMNHRVEITEGILAD
ECAALLSDFFRMRRQEIKAQKKAQSSTD (SEQ ID NO: 317)

ecTadA (H8Y, D108N, N127S)
SEVEFSYEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARNAKTGAAGSLMDVLHHPGMSHRVEITEGILAD
ECAALLSDFFRMRRQEIKAQKKAQSSTD (SEQ ID NO: 318)

ecTadA (H8Y, D108N, N127S, E155D)
SEVEFSYEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARNAKTGAAGSLMDVLHHPGMSHRVEITEGILAD
ECAALLSDFFRMRRQDIKAQKKAQSSTD (SEQ ID NO: 319)

ecTadA (H8Y, D108N, N127S, E155G)

SEVEFSYEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARNAKTGAAGSLMDVLHHPGMSHRVEITEGILAD
ECAALLSDFFRMRRQGIKAQKKAQSSTD (SEQ ID NO: 320)

ecTadA (H8Y, D108N, N127S, E155V)
SEVEFSYEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARNAKTGAAGSLMDVLHHPGMSHRVEITEGILAD
ECAALLSDFFRMRRQVIKAQKKAQSSTD (SEQ ID NO: 321)

ecTadA (A106V, D108N, D147Y, and E155V)
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILAD
ECAALLSYFFRMRRQVIKAQKKAQSSTD (SEQ ID NO:322)

ecTadA (A106V, D108N, D147Y, and E155V)
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILAD
ECAALLSYFFRMRRQVIKAQKKAQSSTD (SEQ ID NO: 407)

ecTadA (L84F, A106V, D108N, H123Y, D147Y, E155V, I156F)
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
ECAALLSYFFRMRRQVFKAQKKAQSSTD (SEQ ID NO: 408)

ecTadA (S2A, I49F, A106V, D108N, D147Y, E155V)
AEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPFGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILAD
ECAALLSYFFRMRRQVIKAQKKAQSSTD (SEQ ID NO: 409)

ecTadA (H8Y, A106T, D108N, N127S, K160S)
SEVEFSYEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGTRNAKTGAAGSLMDVLHHPGMSHRVEITEGILAD
ECAALLSDFFRMRRQEIKAQSKAQSSTD (SEQ ID NO: 410)

ecTadA (R26G, L84F, A106V, R107H, D108N, H123Y, A142N, A143D, D147Y, E155V, I156F)
SEVEFSHEYWMRHALTLAKRAWDEGEVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVHNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
ECNDLLSYFFRMRRQVFKAQKKAQSSTD (SEQ ID NO: 411)

ecTadA (E25G, R26G, L84F, A106V, R107H, D108N, H123Y, A142N, A143D, D147Y, E155V, I156F)
SEVEFSHEYWMRHALTLAKRAWDGGEVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVHNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
ECNDLLSYFFRMRRQVFKAQKKAQSSTD (SEQ ID NO: 412)

ecTadA (E25D, R26G, L84F, A106V, R107K, D108N, H123Y, A142N, A143G, D147Y, E155V, I156F)
SEVEFSHEYWMRHALTLAKRAWDDGEVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVKNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
ECNGLLSYFFRMRRQVFKAQKKAQSSTD (SEQ ID NO: 413)

ecTadA (R26Q, L84F, A106V, D108N, H123Y, A142N, D147Y, E155V, I156F)
SEVEFSHEYWMRHALTLAKRAWDEQEVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
ECNALLSYFFRMRRQVFKAQKKAQSSTD (SEQ ID NO: 414)

ecTadA (E25M, R26G, L84F, A106V, R107P, D108N, H123Y, A142N, A143D, D147Y, E155V, I156F)

SEVEFSHEYWMRHALTLAKRAWDMGEVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVPNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
ECNDLLSYFFRMRRQVFKAQKKAQSSTD (SEQ ID NO: 415)

ecTadA (R26C, L84F, A106V, R107H, D108N, H123Y, A142N , D147Y, E155V, I156F)
SEVEFSHEYWMRHALTLAKRAWDECEVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVHNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
ECNALLSYFFRMRRQVFKAQKKAQSSTD (SEQ ID NO: 416)

ecTadA (L84F, A106V , D108N, H123Y, A142N, A143L, D147Y, E155V, I156F)
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
ECNLLLSYFFRMRRQVFKAQKKAQSSTD (SEQ ID NO: 417)

ecTadA (R26G, L84F, A106V, D108N, H123Y, A142N , D147Y, E155V, I156F)
SEVEFSHEYWMRHALTLAKRAWDEGEVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
ECNALLSYFFRMRRQVFKAQKKAQSSTD (SEQ ID NO: 418)

ecTadA (R51H, L84F, A106V, D108N, H123Y, D147Y, E155V, I156F, K157N)
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGHHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
ECAALLSYFFRMRRQVFNAQKKAQSSTD (SEQ ID NO: 419)

ecTadA (E25A, R26G, L84F, A106V, R107N, D108N, H123Y, A142N, A143E, D147Y, E155V, I156F)
SEVEFSHEYWMRHALTLAKRAWDAGEVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVNNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
ECNELLSYFFRMRRQVFKAQKKAQSSTD (SEQ ID NO: 420)

ecTadA (L84F, A106V, D108N, H123Y, D147Y, E155V, I156F)
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
ECAALLSYFFRMRRQVFKAQKKAQSSTD
(SEQ ID NO: 421)

ecTadA (N37T, P48T, L84F, A106V, D108N, H123Y, D147Y, E155V, I156F)
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHTNRVIGEGWNRTIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
ECAALLSYFFRMRRQVFKAQKKAQSSTD
(SEQ ID NO: 422)

ecTadA (N37S, L84F, A106V, D108N, H123Y, D147Y, E155V, I156F)
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHSNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVM
QNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADE
CAALLSYFFRMRRQVFKAQKKAQSSTD
(SEQ ID NO: 423)

ecTadA (H36L, L84F, A106V, D108N, H123Y, D147Y, E155V, I156F)
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVLNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVM
QNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADE
CAALLSYFFRMRRQVFKAQKKAQSSTD
(SEQ ID NO: 424)

ecTadA (L84F, A106V, D108N, H123Y, S146R, D147Y, E155V, I156F)

SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
ECAALLRYFFRMRRQVFKAQKKAQSSTD
(SEQ ID NO: 425)

ecTadA (H36L, P48L, L84F, A106V, D108N, H123Y, D147Y, E155V, I156F)
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVLNNRVIGEGWNRLIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
ECAALLSYFFRMRRQVFKAQKKAQSSTD
(SEQ ID NO: 426)

ecTadA (H36L, L84F, A106V, D108N, H123Y, D147Y, E155V, K57N, I156F)
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVLNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVM
QNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADE
CAALLSYFFRMRRQVFNAQKKAQSSTD
(SEQ ID NO: 427)

ecTadA (H36L, L84F, A106V, D108N, H123Y, S146C, D147Y, E155V, I156F)
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVLNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVM
QNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADE
CAALLCYFFRMRRQVFKAQKKAQSSTD
(SEQ ID NO: 428)

ecTadA (L84F, A106V, D108N, H123Y, S146R, D147Y, E155V, I156F)
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
ECAALLRYFFRMRRQVFKAQKKAQSSTD
(SEQ ID NO: 429)

ecTadA (N37S, R51H, L84F, A106V, D108N, H123Y, D147Y, E155V, I156F
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHSNRVIGEGWNRPIGHHDPTAHAEIMALRQGGLVM
QNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADE
CAALLSYFFRMRRQVFKAQKKAQSSTD
(SEQ ID NO: 430)

ecTadA (R51L, L84F, A106V, D108N, H123Y, D147Y, E155V, I156F, K157N
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
ECAALLSYFFRMRRQVFNAQKKAQSSTD
(SEQ ID NO: 431)

saTadA (wt)
MGSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAHAEHIAIERAAKVLGS
WRLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCSGSLMNLLQQSNFNHRAIVDKGVLKEACS
TLLTTFFKNLRANKKSTN (SEQ ID NO: 432)

saTadA (D108N)
GSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAHAEHIAIERAAKVLGSW
RLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADNPKGGCSGSLMNLLQQSNFNHRAIVDKGVLKEACST
LLTTFFKNLRANKKSTN (SEQ ID NO: 433)

saTadA (D107A_D108N)
GSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAHAEHIAIERAAKVLGSW
RLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGAANPKGGCSGSLMNLLQQSNFNHRAIVDKGVLKEACST
LLTTFFKNLRANKKSTN (SEQ ID NO: 434)

saTadA (G26P_D107A_D108N)
GSHMTNDIYFMTLAIEEAKKAAQLPEVPIGAIITKDDEVIARAHNLRETLQQPTAHAEHIAIERAAKVLGSW
RLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGAANPKGGCSGSLMNLLQQSNFNHRAIVDKGVLKEACST
LLTTFFKNLRANKKSTN (SEQ ID NO: 435)

saTadA (G26P_D107A_D108N_S142A)
GSHMTNDIYFMTLAIEEAKKAAQLPEVPIGAIITKDDEVIARAHNLRETLQQPTAHAEHIAIERAAKVLGSW
RLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGAANPKGGCSGSLMNLLQQSNFNHRAIVDKGVLKEACAT
LLTTFFKNLRANKKSTN (SEQ ID NO: 436)

saTadA (D107A_D108N_S142A)
GSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAHAEHIAIERAAKVLGSW
RLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGAANPKGGCSGSLMNLLQQSNFNHRAIVDKGVLKEACAT
LLTTFFKNLRANKKSTN (SEQ ID NO: 437)

ecTadA (P48S)
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRSIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILAD
ECAALLSDFFRMRRQEIKAQKKAQSSTD (SEQ ID NO: 438)

ecTadA (P48T)
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRTIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILAD
ECAALLSDFFRMRRQEIKAQKKAQSSTD (SEQ ID NO: 439)

ecTadA (P48A)
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRAIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILAD
ECAALLSDFFRMRRQEIKAQKKAQSSTD (SEQ ID NO: 440)

ecTadA (A142N)
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILAD
ECNALLSDFFRMRRQEIKAQKKAQSSTD (SEQ ID NO: 441)

ecTadA (W23R)
SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVM
QNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADE
CAALLSDFFRMRRQEIKAQKKAQSSTD (SEQ ID NO: 442)

ecTadA (W23L)
SEVEFSHEYWMRHALTLAKRALDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVM
QNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADE
CAALLSDFFRMRRQEIKAQKKAQSSTD (SEQ ID NO: 443)

ecTadA (R152P)
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILAD
ECAALLSDFFRMPRQEIKAQKKAQSSTD (SEQ ID NO: 444)

ecTadA (R152H)
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILAD
ECAALLSDFFRMHRQEIKAQKKAQSSTD (SEQ ID NO: 445)

ecTadA (L84F, A106V, D108N, H123Y, D147Y, E155V, I156F)
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
ECAALLSYFFRMRRQVFKAQKKAQSSTD (SEQ ID NO: 446)

ecTadA (H36L, R51L, L84F, A106V, D108N, H123Y, S146C, D147Y, E155V, I156F, K157N)
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVLNNRVIGEGWNRPIGLHDPTAHAEIMALRQGGLVM
QNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADE
CAALLCYFFRMRRQVFNAQKKAQSSTD (SEQ ID NO: 447)

ecTadA (H36L, P48S, R51L, L84F, A106V, D108N, H123Y, S146C, D147Y, E155V, I156F , K157N)
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVLNNRVIGEGWNRSIGLHDPTAHAEIMALRQGGLVM
QNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADE
CAALLCYFFRMRRQVFNAQKKAQSSTD (SEQ ID NO: 448)

ecTadA (H36L, P48A, R51L, L84F, A106V, D108N, H123Y, S146C, D147Y, E155V, I156F , K157N)
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
ECAALLCYFFRMRRQVFNAQKKAQSSTD (SEQ ID NO: 449)

ecTadA (W23L, H36L, P48A, R51L, L84F, A106V, D108N, H123Y, S146C, D147Y, R152P, E155V, I156F,
K157N)
SEVEFSHEYWMRHALTLAKRALDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVM
QNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADE
CAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 450)
ecTadA (W23R, H36L, P48A, R51L, L84F, A106V, D108N, H123Y, S146C, D147Y, R152P, E155V, I156F,
K157N)
SEVEFSHEYWMRHALTLAKRALDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVM
QNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADE
CAALLCYFFRMPRQVFNAQKKAQSSTD
(SEQ ID NO: 479)

Staphylococcus aureus TadA:
MGSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAHAEHIAIERAAKVLGS
WRLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCSGSLMNLLQQSNFNHRAIVDKGVLKEACS
TLLTTFFKNLRANKKSTN (SEQ ID NO: 451)

Bacillus subtilis TadA:
MTQDELYMKEAIKEAKKAEEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEMLVID
EACKALGTWRLEGATLYVTLEPCPMCAGAVVLSRVEKVVFGAFDPKGGCSGTLMNLLQEERFNHQAEVV
SGVLEEECGGMLSAFFRELRKKKKAARKNLSE (SEQ ID NO: 452)

Salmonella typhimurium (*S. typhimurium*) TadA:
MPPAFITGVTSLSDVELDHEYWMRHALTLAKRAWDEREVPVGAVLVHNHRVIGEGWNRPIGRHDPTAHA
EIMALRQGGLVLQNYRLLDTTLYVTLEPCVMCAGAMVHSRIGRVVFGARDAKTGAAGSLIDVLHHPGMN
HRVEIIEGVLRDECATLLSDFFRMRRQEIKALKKADRAEGAGPAV (SEQ ID NO: 453)

Shewanella putrefaciens (*S. putrefaciens*) TadA:
MDEYWMQVAMQMAEKAEAAGEVPVGAVLVKDGQQIATGYNLSISQHDPTAHAEILCLRSAGKKLENYR
LLDATLYITLEPCAMCAGAMVHSRIARVVYGARDEKTGAAGTVVNLLQHPAFNHQVEVTSGVLAEACSA
QLSRFFKRRRDEKKALKLAQRAQQGIE (SEQ ID NO: 454)

Haemophilus influenzae F3031 (*H. influenzae*) TadA:
MDAAKVRSEFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQSDPTAHAEIIALRNG
AKNIQNYRLLNSTLYVTLEPCTMCAGAILHSRIKRLVFGASDYKTGAIGSRFHFFDDYKMNHTLEITSGVLA
EECSQKLSTFFQKRREEKKIEKALLKSLSDK (SEQ ID NO: 455)

Caulobacter crescentus (C. crescentus) TadA:
MRTDESEDQDHRMMRLALDAARAAAEAGETPVGAVILDPSTGEVIATAGNGPIAAHDPTAHAEIAAMRA
AAAKLGNYRLTDLTLVVTLEPCAMCAGAISHARIGRVVFGADDPKGGAVVHGPKFFAQPTCHWRPEVTG
GVLADESADLLRGFFRARRKAKI (SEQ ID NO: 456)

Geobacter sulfurreducens (*G. sulfurreducens*) TadA:
MSSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRGHNLREGSNDPSAHAEMIAIRQA
ARRSANWRLTGATLYVTLEPCLMCMGAIILARLERVVFGCYDPKGGAAGSLYDLSADPRLNHQVRLSPGV
CQEECGTMLSDFFRDLRRRKKAKATPALFIDERKVPPEP (SEQ ID NO: 457)

[00159] Non-limiting examples of A to G nucleobase editors are provided.

ecTadA(wt)-XTEN-nCas9-NLS (SEQ ID NO: 323)

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILA
DECAALLSDFFRMRRQEIKAQKKAQSSTDSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVP
SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPL
SASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

ecTadA(D108N)-XTEN-nCas9-NLS: (mammalian construct, active on DNA, A to G editing,

SEQ ID NO: 324)

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARNAKTGAAGSLMDVLHHPGMNHRVEITEGILA
DECAALLSDFFRMRRQEIKAQKKAQSSTDSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVP
SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPL
SASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL

VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

ecTadA(D108G)-XTEN-nCas9-NLS: (mammalian construct, active on DNA, A to G editing,

SEQ ID NO:325)

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARGAKTGAAGSLMDVLHHPGMNHRVEITEGILA
DECAALLSDFFRMRRQEIKAQKKAQSSTDSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVP
SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPL
SASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

ecTadA(D108V)-XTEN-nCas9-NLS: (mammalian construct, active on DNA, A to G editing,

SEQ ID NO: 326)

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARVAKTGAAGSLMDVLHHPGMNHRVEITEGILA
DECAALLSDFFRMRRQEIKAQKKAQSSTDSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVP
SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPL
SASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN

GRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

ecTadA(D108N)-XTEN-nCas9-UGI-NLS (BE3 analog of A to G editor, SEQ ID NO: 327)

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARNAKTGAAGSLMDVLHHPGMNHRVEITEGILA
DECAALLSDFFRMRRQEIKAQKKAQSSTDSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVP
SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPL
SASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQE
SILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKK
RKV

ecTadA(D108G)-XTEN-nCas9-UGI-NLS (BE3 analog of A to G editor, SEQ ID NO: 328):

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARGAKTGAAGSLMDVLHHPGMNHRVEITEGILA
DECAALLSDFFRMRRQEIKAQKKAQSSTDSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVP
SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPL
SASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI

GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQE
SILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKK
RKV

ecTadA(D108V)-XTEN-nCas9-UGI-NLS (BE3 analog of A to G editor, SEQ ID NO: 329):

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARVAKTGAAGSLMDVLHHPGMNHRVEITEGILA
DECAALLSDFFRMRRQEIKAQKKAQSSTDSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVP
SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPL
SASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQE
SILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKK
RKV

ecTadA(D108N)-XTEN-dCas9-UGI-NLS (mammalian cells, BE2 analog of A to G editor, SEQ ID NO: 330):

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARNAKTGAAGSLMDVLHHPGMNHRVEITEGILA
DECAALLSDFFRMRRQEIKAQKKAQSSTDSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVP
SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPL
SASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK

SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQE
SILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKK
RKV

ecTadA(D108G)-XTEN-dCas9-UGI-NLS (mammalian cells, BE2 analog of A to G editor, SEQ ID NO: 331):

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARGAKTGAAGSLMDVLHHPGMNHRVEITEGILA
DECAALLSDFFRMRRQEIKAQKKAQSSTDSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVP
SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPL
SASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQE
SILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKK
RKV

ecTadA(D108V)-XTEN-dCas9-UGI-NLS (mammalian cells, BE2 analog of A to G editor, SEQ ID NO: 332):

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARVAKTGAAGSLMDVLHHPGMNHRVEITEGILA
DECAALLSDFFRMRRQEIKAQKKAQSSTDSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVP
SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPL
SASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL

VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQE
SILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKM
LSGGSPKKKRKV

ecTadA(D108N)-XTEN-nCas9-AAG(E125Q)-NLS – cat. alkyladenosine glycosylase (SEQ ID NO: 333)

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARNAKTGAAGSLMDVLHHPGMNHRVEITEGILA
DECAALLSDFFRMRRQEIKAQKKAQSSTDSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVP
SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPL
SASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSKGHLTRLGLEFFDQPAVP
LARAFLGQVLVRRLPNGTELRGRIVETQAYLGPEDEAAHSRGGRQTPRNRGMFMKPGTLYVYIIYGMYFC
MNISSQGDGACVLLRALEPLEGLETMRQLRSTLRKGTASRVLKDRELCSGPSKLCQALAINKSFDQRDLAQ
DEAVWLERGPLEPSEPAVVAAARVGVGHAGEWARKPLRFYVRGSPWVSVVDRVAEQDTQASGGSPKKK
RKV

ecTadA(D108G)-XTEN-nCas9-AAG(E125Q)-NLS – cat. alkyladenosine glycosylase (SEQ ID NO: 334)

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARGAKTGAAGSLMDVLHHPGMNHRVEITEGILA
DECAALLSDFFRMRRQEIKAQKKAQSSTDSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVP
SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPL
SASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL

LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSKGHLTRLGLEFFDQPAVP
LARAFLGQVLVRRLPNGTELRGRIVETQAYLGPEDEAAHSRGGRQTPRNRGMFMKPGTLYVYIIYGMYFC
MNISSQGDGACVLLRALEPLEGLETMRQLRSTLRKGTASRVLKDRELCSGPSKLCQALAINKSFDQRDLAQ
DEAVWLERGPLEPSEPAVVAAARVGVGHAGEWARKPLRFYVRGSPWVSVVDRVAEQDTQASGGSPKKK
RKV

ecTadA(D108V)-XTEN-nCas9-AAG(E125Q)-NLS – cat. alkyladenosine glycosylase (SEQ ID NO: 335)

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARVAKTGAAGSLMDVLHHPGMNHRVEITEGILA
DECAALLSDFFRMRRQEIKAQKKAQSSTDSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVP
SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPL
SASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSKGHLTRLGLEFFDQPAVP
LARAFLGQVLVRRLPNGTELRGRIVETQAYLGPEDEAAHSRGGRQTPRNRGMFMKPGTLYVYIIYGMYFC
MNISSQGDGACVLLRALEPLEGLETMRQLRSTLRKGTASRVLKDRELCSGPSKLCQALAINKSFDQRDLAQ
DEAVWLERGPLEPSEPAVVAAARVGVGHAGEWARKPLRFYVRGSPWVSVVDRVAEQDTQASGGSPKKK
RKV

ecTadA(D108N)-XTEN-nCas9-EndoV(D35A)-NLS: contains cat. endonuclease V (SEQ ID NO: 336)

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARNAKTGAAGSLMDVLHHPGMNHRVEITEGILA
DECAALLSDFFRMRRQEIKAQKKAQSSTDSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVP
SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPL
SASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSDLASLRAQQIELASSVIRE
DRLDKDPPDLIAGAAVGFEQGGEVTRAAMVLLKYPSLELVEYKVARIATTMPYIPGFLSFREYPALLAAWE
MLSQKPDLVFVDGHGISHPRRLGVASHFGLLVDVPTIGVAKKRLCGKFEPLSSEPGALAPLMDKGEQLAW
VWRSKARCNPLFIATGHRVSVDSALAWVQRCMKGYRLPEPTRWADAVASERPAFVRYTANQPSGGSPKK
KRKV

ecTadA(D108G)-XTEN-nCas9-EndoV (D35A)-NLS: contains cat. endonuclease V (SEQ ID NO: 337)

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARGAKTGAAGSLMDVLHHPGMNHRVEITEGILA
DECAALLSDFFRMRRQEIKAQKKAQSSTDSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVP
SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPL
SASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSDLASLRAQQIELASSVIRE
DRLDKDPPDLIAGAAVGFEQGGEVTRAAMVLLKYPSLELVEYKVARIATTMPYIPGFLSFREYPALLAAWE
MLSQKPDLVFVDGHGISHPRRLGVASHFGLLVDVPTIGVAKKRLCGKFEPLSSEPGALAPLMDKGEQLAW

VWRSKARCNPLFIATGHRVSVDSALAWVQRCMKGYRLPEPTRWADAVASERPAFVRYTANQPSGGSPKK
KRKV

ecTadA(D108V)-XTEN-nCas9-EndoV(D35A)-NLS: contains cat. endonuclease V (SEQ ID NO: 338)

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARVAKTGAAGSLMDVLHHPGMNHRVEITEGILA
DECAALLSDFFRMRRQEIKAQKKAQSSTDSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVP
SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPL
SASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSDLASLRAQQIELASSVIRE
DRLDKDPPDLIAGAAVGFEQGGEVTRAAMVLLKYPSLELVEYKVARIATTMPYIPGFLSFREYPALLAAWE
MLSQKPDLVFVDGHGISHPRRLGVASHFGLLVDVPTIGVAKKRLCGKFEPLSSEPGALAPLMDKGEQLAW
VWRSKARCNPLFIATGHRVSVDSALAWVQRCMKGYRLPEPTRWADAVASERPAFVRYTANQPSGGSPKK
KRKV

Variant resulting from first round of evolution (in bacteria) ecTadA(H8Y_D108N_N127S)-XTEN-dCas9 (SEQ ID NO: 339)

MSEVEFSYEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARNAKTGAAGSLMDVLHHPGMSHRVEITEGILA
DECAALLSDFFRMRRQEIKAQKKAQSSTDSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVP
SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPL
SASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI

GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

Enriched variants from second round of evolution (in bacteria) ecTadA

(H8Y_D108N_N127S_E155X)-XTEN-dCas9; X=D, G or V (SEQ ID NO: 340)

MSEVEFSYEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARNAKTGAAGSLMDVLHHPGMSHRVEITEGILA
DECAALLSDFFRMRRQXIKAQKKAQSSTDSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVP
SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPL
SASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

pNMG-160: ecTadA(D108N)-XTEN-nCas9-GGS-AAG*(E125Q)-GGS-NLS (SEQ ID NO:

341)

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARNAKTGAAGSLMDVLHHPGMNHRVEITEGILA
DECAALLSDFFRMRRQEIKAQKKAQSSTDSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVP
SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPL
SASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ

TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSKGHLTRLGLEFFDQPAVPL
ARAFLGQVLVRRLPNGTELRGRIVETQAYLGPEDEAAHSRGGRQTPRNRGMFMKPGTLYVYIIYGMYFCM
NISSQGDGACVLLRALEPLEGLETMRQLRSTLRKGTASRVLKDRELCSGPSKLCQALAINKSFDQRDLAQD
EAVWLERGPLEPSEPAVVAAARVGVGHAGEWARKPLRFYVRGSPWVSVVDRVAEQDTQAGGSPKKKRK
V

pNMG-161: ecTadA(D108N)-XTEN-nCas9-GGS-EndoV*(D35A)-GGS-NLS (SEQ ID NO: 342)

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARNAKTGAAGSLMDVLHHPGMNHRVEITEGILA
DECAALLSDFFRMRRQEIKAQKKAQSSTDSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVP
SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPL
SASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSDLASLRAQQIELASSVIRE
DRLDKDPPDLIAGAAVGFEQGGEVTRAAMVLLKYPSLELVEYKVARIATTMPYIPGFLSFREYPALLAAWE
MLSQKPDLVFVDGHGISHPRRLGVASHFGLLVDVPTIGVAKKRLCGKFEPLSSEPGALAPLMDKGEQLAW
VWRSKARCNPLFIATGHRVSVDSALAWVQRCMKGYRLPEPTRWADAVASERPAFVRYTANQPGGSPKKK
RKV

pNMG-371: ecTadA(L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)-SGGS-SGGS-XTEN-SGGS-SGGS- ecTadA(L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)-SGGS-SGGS-XTEN-SGGS-SGGS-nCas9-SGGS- NLS (SEQ ID NO: 458)

SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHA
EIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGS
LMDVLHYPGMNHRVEITEGILADECAALLSYFFRMRRQVFKAQKKAQSSTDSGGSSGG
SSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLV
HNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGA
MIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLSYFFRM
RRQVFKAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGT
NSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRY
TRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY

PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFK
SNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEIT
KAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFL
KDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIE
RMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL
LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE
NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR
DKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP
AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE
LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD
DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK
VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL
VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHK
HYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

pNMG-616 amino acid sequence: ecTadA(wild type)-(SGGS)2-XTEN-(SGGS)2-
ecTadA(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E1
55V_I156F _K157N)-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS (SEQ ID NO: 459)
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAG
SLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGG
SSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRALDEREVPVGAVLV
LNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGA
MIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRM
PRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTN
SVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYP
TIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYN
QLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKS
NFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIER
MTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLL
FKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN
EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD
KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKEL
GSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD

DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK
VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL
VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHK
HYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

pNMG-624 amino acid sequence: ecTadA(wild type)-32 a.a. linker-
ecTadA(W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E1
55V_I156F _K157N)-24 a.a. linker_nCas9_SGGS_NLS (SEQ ID NO: 460)
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAG
SLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGG
SSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLV
LNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGA
MIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRM
PRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVI
TDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRIC
YLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED
AKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASM
IKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE
KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIE
KILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKN
LPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT
VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVL
TLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT
VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEH
PVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL
TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA
GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK
VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKA
TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQV
NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEK
GKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKR
MLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIE
QISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID
RKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

pNMG-476 amino acid sequence (evolution #3 hetero dimer, wt TadA + TadA evo #3
mutations): ecTadA(wild-type)-(SGGS)2-XTEN-(SGGS)2-

ecTadA(L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS (SEQ ID NO: 461)
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAG
SLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGG
SSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLV
HNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGA
MIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLSYFFRM
RRQVFKAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGT
NSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRY
TRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY
PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFK
SNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEIT
KAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFL
KDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIE
RMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL
LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE
NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR
DKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP
AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE
LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD
DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK
VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL
VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHK
HYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

pNMG-477 amino acid sequence: ecTadA(wild-type)-(SGGS)2-XTEN-(SGGS)2-ecTadA(H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N)-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS (SEQ ID NO: 462)
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAG
SLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGG
SSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLV
LNNRVIGEGWNRPIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGA
MIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRM
RRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGT
NSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRY
TRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY
PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY

NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFK
SNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEIT
KAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFL
KDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIE
RMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL
LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE
NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR
DKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP
AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE
LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD
DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK
VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL
VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHK
HYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

pNMG-558 amino acid sequence: ecTadA(wild-type)- 32 a.a. linker-ecTadA(H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N)-24 a.a. linker_nCas9_SGGS_NLS (SEQ ID NO: 463)
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAG
SLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGG
SSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLV
LNNRVIGEGWNRPIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGA
MIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRM
RRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVI
TDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRIC
YLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED
AKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASM
IKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE
KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIE
KILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKN
LPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT
VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVL
TLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT
VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEH
PVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL

TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA
GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK
VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKA
TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQV
NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEK
GKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKR
MLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIE
QISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID
RKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

pNMG-576 amino acid sequence: ecTadA(wild-type)-(SGGS)2-XTEN-(SGGS)2-
ecTadA(H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F
_K157N)-(SGGS)2-XTEN-(SGGS)2_nCas9_GGS_NLS (SEQ ID NO: 464)
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAG
SLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGG
SSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLV
LNNRVIGEGWNRSIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGA
MIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRM
RRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGT
NSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRY
TRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY
PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFK
SNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEIT
KAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFL
KDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIE
RMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL
LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE
NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR
DKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP
AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE
LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD
DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK
VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL
VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHK
HYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

pNMG-577 amino acid sequence: ecTadA(wild-type)-(SGGS)2-XTEN-(SGGS)2-ecTadA(H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F _K157N)-(SGGS)2-XTEN-(SGGS)2_nCas9_GGS_NLS (SEQ ID NO: 465)

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVLNNRVIGEGWNRSIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECNALLCYFFRMRRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

pNMG-586 amino acid sequence: ecTadA(wild-type)-(SGGS)2-XTEN-(SGGS)2-ecTadA(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N)-(SGGS)2-XTEN-(SGGS)2_nCas9_GGS_NLS (SEQ ID NO: 466)

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMRRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY

PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFK
SNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEIT
KAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFL
KDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIE
RMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL
LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE
NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR
DKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP
AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE
LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD
DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK
VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL
VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHK
HYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

pNMG-588 amino acid sequence: ecTadA(wild-type)-(SGGS)2-XTEN-(SGGS)2-
ecTadA(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I
156F _K157N)-(SGGS)2-XTEN-(SGGS)2_nCas9_GGS_NLS (SEQ ID NO: 467)
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAG
SLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGG
SSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLV
LNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGA
MIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECNALLCYFFRM
RRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGT
NSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRY
TRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY
PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFK
SNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEIT
KAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFL
KDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIE
RMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL
LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE
NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR
DKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP
AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE

LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD
DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK
VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL
VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHK
HYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

pNMG-620 amino acid sequence: ecTadA(wild-type)-(SGGS)2-XTEN-(SGGS)2-
ecTadA(W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E1
55V_I156F _K157N)-(SGGS)2-XTEN-(SGGS)2_nCas9_GGS_NLS (SEQ ID NO: 468)
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAG
SLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGG
SSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLV
LNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGA
MIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRM
PRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTN
SVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYP
TIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYN
QLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKS
NFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIER
MTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLL
FKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN
EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD
KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKEL
GSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD
DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK
VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL
VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHK
HYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

pNMG-617 amino acid sequence: ecTadA(wild-type)-(SGGS)2-XTEN-(SGGS)2-ecTadA(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_E155V_I156F _K157N)-(SGGS)2-XTEN-(SGGS)2_nCas9_GGS_NLS (SEQ ID NO: 469)

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAG
SLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGG
SSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRALDEREVPVGAVLV
LNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGA
MIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECNALLCYFFRM
RRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGT
NSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRY
TRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY
PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFK
SNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEIT
KAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFL
KDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIE
RMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL
LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE
NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR
DKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP
AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE
LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD
DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK
VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL
VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHK
HYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

pNMG-618 amino acid sequence: ecTadA(wild-type)-(SGGS)2-XTEN-(SGGS)2-ecTadA(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_R152P_E155V_I156F _K157N)-(SGGS)2-XTEN-(SGGS)2_nCas9_GGS_NLS (SEQ ID NO: 470)

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAG
SLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGG
SSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRALDEREVPVGAVLV
LNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGA
MIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECNALLCYFFRM
PRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTN
SVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT

RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYP
TIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYN
QLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKS
NFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIER
MTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLL
FKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN
EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD
KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKEL
GSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD
DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK
VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL
VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHK
HYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

pNMG-620 amino acid sequence: ecTadA(wild-type)-(SGGS)2-XTEN-(SGGS)2-
ecTadA(W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E1
55V_I156F _K157N)-(SGGS)2-XTEN-(SGGS)2_nCas9_GGS_NLS (SEQ ID NO: 471)
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAG
SLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGG
SSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLV
LNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGA
MIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRM
PRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTN
SVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYP
TIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYN
QLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKS
NFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIER
MTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLL
FKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN
EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD
KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA

IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKEL
GSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD
DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK
VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL
VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHK
HYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

pNMG-621 amino acid sequence: ecTadA(wild-type)- 32 a.a. linker-
ecTadA(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I1
56F _K157N)- 24 a.a. linker_nCas9_GGS_NLS (SEQ ID NO: 472)
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAG
SLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGG
SSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLV
LNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGA
MIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRM
PRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVI
TDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRIC
YLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED
AKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASM
IKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE
KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIE
KILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKN
LPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT
VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVL
TLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT
VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEH
PVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL
TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA
GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK
VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKA
TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQV
NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEK
GKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKR
MLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIE
QISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID
RKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

pNMG-622 amino acid sequence: ecTadA(wild-type)- 32 a.a. linker-ecTadA(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_R152P_E155V_I156F _K157N)- 24 a.a. linker_nCas9_GGS_NLS (SEQ ID NO: 473)

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAG
SLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGG
SSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLV
LNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGA
MIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECNALLCYFFRM
PRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVI
TDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRIC
YLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED
AKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASM
IKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE
KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIE
KILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKN
LPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT
VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVL
TLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT
VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEH
PVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL
TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA
GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK
VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKA
TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQV
NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEK
GKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKR
MLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIE
QISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID
RKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

pNMG-623 amino acid sequence: ecTadA(wild-type)- 32 a.a. linker-ecTadA(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F _K157N)- 24 a.a. linker_nCas9_GGS_NLS (SEQ ID NO: 474)

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAG
SLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGG
SSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRALDEREVPVGAVLV
LNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGA
MIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRM

PRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVI
TDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRIC
YLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED
AKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASM
IKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE
KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIE
KILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKN
LPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT
VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVL
TLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT
VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEH
PVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL
TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA
GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK
VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKA
TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQV
NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEK
GKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKR
MLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIE
QISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID
RKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

ABE6.3 ecTadA(wild-type)-(SGGS)2-XTEN-(SGGS)2-
ecTadA(H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K1
57N)-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS (SEQ ID NO: 475)
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAG
SLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGG
SSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLV
LNNRVIGEGWNRSIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGA
MIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRM
RRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGT
NSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRY
TRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY
PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFK
SNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEIT
KAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFL
KDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIE
RMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL
LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE
NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR

DKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP
AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE
LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD
DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK
VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL
VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHK
HYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV*

ABE7.8 ecTadA(wild-type)-(SGGS)2-XTEN-(SGGS)2-
ecTadA(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E
155V_I156F_K157N)-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS (SEQ ID NO: 476)
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAG
SLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGG
SSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRALDEREVPVGAVLV
LNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGA
MIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECNALLCYFFRM
RRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGT
NSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRY
TRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY
PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFK
SNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEIT
KAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFL
KDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIE
RMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL
LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE
NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR
DKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP
AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE
LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD
DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK
VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL
VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHK
HYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV*

ABE7.9 ecTadA(wild-type)-(SGGS)2-XTEN-(SGGS)2-ecTadA(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_R152P¬_E155V_I156F_K157N)-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS (SEQ ID NO: 477)

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAG
SLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGG
SSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRALDEREVPVGAVLV
LNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGA
MIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECNALLCYFFRM
PRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTN
SVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYP
TIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYN
QLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKS
NFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIER
MTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLL
FKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN
EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD
KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKEL
GSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD
DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK
VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL
VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHK
HYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV*

ABE7.10 ecTadA(wild-type)-(SGGS)2-XTEN-(SGGS)2-ecTadA(W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P¬_E155V_I156F_K157N)-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAG
SLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGG
SSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLV
LNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGA
MIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRM

PRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTN
SVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYP
TIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYN
QLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKS
NFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIER
MTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLL
FKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN
EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD
KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKEL
GSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD
DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK
VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL
VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHK
HYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV*

ABE6.4: ecTadA(wild-type)-(SGGS)2-XTEN-(SGGS)2-
ecTadA(H36L_P48S_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I1
56F_K157N)-(SGGS)2-XTEN-(SGGS)2_nCas9_SGGS_NLS (SEQ ID NO: 480)
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAG
SLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGG
SSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLV
LNNRVIGEGWNRSIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGA
MIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECNALLCYFFRM
RRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGT
NSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRY
TRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY
PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFK
SNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEIT
KAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFL
KDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIE
RMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL
LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE
NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR

DKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP
AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE
LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD
DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK
VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL
VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHK
HYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

Example 2: AAV Delivery of Split Nucleobase Editor

[00160] Thisstudy was designed to show that a nucleobase editor may be delivered by recombinant AAV (rAAV) in two sections, which may be joined to form a complete and active nucleobase editor in cells via protein splicing. Different elements of the rAAV constructs were tested for optimized nucleobase editor expression and activity.

[00161] Recombinant AAV (rAAV) is widely used for transgene delivery. Transgenes were inserted into the AAV genome between the inverted terminal repeat (ITR) sequences and packaged into AAV viral particles, which are used to transduce a host cell (*e.g.*, mammalian cell, human cell). However, there is a limitation on the size of the transgene that may be packaged into rAAV, typically approximately 4.9 kilobases. Nucleic acids encoding a nucleobase editor (*e.g.*, cytosine deaminase-dCas9-UGI) typically exceed the packaging limit of rAAV. As described herein, the nucleic acids encoding a nucleobase editor were split (see Figure 1A), and each section was packaged into a separate rAAV particle. The two sections of the nucleobase editor were delivered to the cells and can be ligated to form a complete nucleobase editor via protein splicing (*e.g.*, mediated by an intein, such as the DnaE intein; see Figure 1C). The ligated, complete nucleobase editor was active in editing target bases (see Figure 1B). The rAAV constructs encoding the split nucleobase editors were tested in different cell lines, *e.g.*, U118 and HEK293T, and are active in editing the target base (see Figures 3A-3B and Figures 5A-5B).

[00162] Different transcriptional terminators and nuclear localization signals (NLS) were tested in the rAAV constructs to optimize the expression and activity of the nucleobase editors (see Figures 4, 6, and 7).

Example 3: Editing of DNMT1 gene in mouse neuron using AAV encoded split nucleobase editor

[00163] This study was designed to test the base editing activity of an AAV encoded split nucleobase editor *in vivo*. A split nucleobase editor as shown in Figure 1A was used. The amino acid sequence of the linker between the dCas9 domain and the deaminase domain is SGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID NO: 384). A guide RNA targeting a well-characterized site in the *DNMT1* gene was selected. It was expected that the cells would be able to tolerate the editing. These experiments aim to determine whether AAV encoded split base editor can edit the locus *in vitro* or *in vivo* in several cell types including primary neurons.

[00164] In one experiment, AAV vectors encoding the split nucleobase editor and a guide RNA targeting *DNMT1* were used to transduce dissociated mouse cortical neurons, two days after the cortical neurons were isolated and cultured. The neurons were harvested 16 days post transduction and the *DNMT1* gene was sequenced (Figure 8A) to determine editing efficiency as well as off-target effects. An editing efficiency of 17.34% (C to T editing, darker grey in Figure 8B) was detected, while only 0.82% of undesired editing (C to G or C to A change, lighter grey in Figure 8B) was detected.

[00165] In another experiment, cultured mouse Neuro-2 cells were either transduced with AAV vectors encoding the split nucleobase editor and a guide RNA targeting *DNMT1*, or transfected with lipid-encapsulated DNA encoding the nucleobase editor and guide RNA, allowing direct comparison of editing efficiency using different delivery methods of the nucleobase editor (Figure 9A). An editing efficiency of 5.96% (C to T editing, dark grey in Figure 9B) was observed for AAV encoded split nucleobase editor, while an editing efficiency of 27.3% (C to T editing, dark grey in Figure 9B) was observed for lipid-transfected DNA encoded nucleobase editor. The amount of undesired products was 0.15% for AAV encoded split nucleobase editor and 1.3% for lipid-transfected DNA encoded nucleobase editor (C to G or C to A change, lighter grey in Figure 9B).

EQUIVALENTS AND SCOPE

[00166] In the claims articles such as "a," "an," and "the" may mean one or more than one unless indicated to the contrary or otherwise evident from the context. Claims or descriptions that include "or" between one or more members of a group are considered satisfied if one, more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process unless indicated to the contrary or otherwise evident from the context.

The invention includes embodiments in which exactly one member of the group is present in, employed in, or otherwise relevant to a given product or process. The invention includes embodiments in which more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process.

[00167] Furthermore, the invention encompasses all variations, combinations, and permutations in which one or more limitations, elements, clauses, and descriptive terms from one or more of the listed claims is introduced into another claim. For example, any claim that is dependent on another claim can be modified to include one or more limitations found in any other claim that is dependent on the same base claim. Where elements are presented as lists, *e.g.*, in Markush group format, each subgroup of the elements is also disclosed, and any element(s) can be removed from the group. It should it be understood that, in general, where the invention, or aspects of the invention, is/are referred to as comprising particular elements and/or features, certain embodiments of the invention or aspects of the invention consist, or consist essentially of, such elements and/or features. For purposes of simplicity, those embodiments have not been specifically set forth *in haec verba* herein.

[00168] It is also noted that the terms "comprising" and "containing" are intended to be open and permits the inclusion of additional elements or steps. Where ranges are given, endpoints are included. Furthermore, unless otherwise indicated or otherwise evident from the context and understanding of one of ordinary skill in the art, values that are expressed as ranges can assume any specific value or sub-range within the stated ranges in different embodiments of the invention, to the tenth of the unit of the lower limit of the range, unless the context clearly dictates otherwise.

[00169] This application refers to various issued patents, published patent applications, journal articles, and other publications, all of which are incorporated herein by reference. If there is a conflict between any of the incorporated references and the instant specification, the specification shall control. In addition, any particular embodiment of the present invention that falls within the prior art may be explicitly excluded from any one or more of the claims. Because such embodiments are deemed to be known to one of ordinary skill in the art, they may be excluded even if the exclusion is not set forth explicitly herein. Any particular embodiment of the invention can be excluded from any claim, for any reason, whether or not related to the existence of prior art.

[00170] Those skilled in the art will recognize or be able to ascertain using no more than routine experimentation many equivalents to the specific embodiments described herein. The scope of the present embodiments described herein is not intended to be limited to the above Description, but rather is as set forth in the appended claims. Those of ordinary skill in the art will appreciate that various changes and modifications to this description may be made without departing from the spirit or scope of the present invention, as defined in the following claims.

CLAIMS

What is claimed is:

1. A composition comprising:

(i) a first nucleotide sequence encoding a N-terminal portion of a Cas9 protein fused at its C-terminus to an intein-N;

(ii) a second nucleotide sequence encoding an intein-C fused to the N-terminus of a C-terminal portion of the Cas9 protein;

wherein the first nucleotide sequence or second nucleotide sequence is operably linked to a nucleotide sequence encoding at least one bipartite nuclear localization signal;

wherein the N-terminal portion of the Cas9 protein comprises a portion of any one of SEQ ID NOs: 1-275 and 394-397 that corresponds to amino acids 1-573 or 1-637 of SEQ ID NO: 1;

wherein the C-terminal portion of the Cas9 protein comprises a portion of any one of SEQ ID NOs: 1-275 and 394-397 that corresponds to amino acids 574-1368 or 638-1368 of SEQ ID NO: 1; and

wherein the intein-C comprises the amino acid sequence as set forth in any one of SEQ ID NOs: 353 and 357.

2. The composition of claim 1, wherein the intein-N comprises the amino acid sequence as set forth in any one of SEQ ID NOs: 351 and 355.

3. The composition of claim 1 or 2, wherein the first nucleotide sequence or the second nucleotide sequence further comprises a nucleotide encoding a guide RNA (gRNA) operably linked to a promoter.

4. The composition of any one of claims 1-3, wherein the first nucleotide sequence and/or the second nucleotide sequence further comprises a transcriptional terminator.

5. The composition of claim 4, wherein the transcriptional terminator is the transcriptional terminator from a bGH gene, hGH gene, or SV40 gene.

6. The composition of claim 5, wherein the transcriptional terminator is the transcriptional terminator from a bGH gene.

7. The composition of any one of claims 1-6, wherein the first nucleotide sequence and/or the second nucleotide sequence further comprises a woodchuck hepatitis posttranscriptional regulatory element (WPRE) inserted 5′ of the transcriptional terminator.

8. The composition of any one of claims 1-7, wherein the bipartite nuclear localization signal comprises an amino acid sequence selected from the group consisting of: KRPAATKKAGQAKKKK (SEQ ID NO: 344), KKTELQTTNAENKTKKL (SEQ ID NO: 345), KRGINDRNFWRGENGRKTR (SEQ ID NO: 346), and RKSGKIAAIVVKRPRK (SEQ ID NO: 347).

9. The composition of any one of claims 1-8, wherein the bipartite nuclear localization signal comprises the amino acid sequence as set forth in SEQ ID NO: 344.

10. The composition of any one of claims 1-9, wherein the Cas9 protein is a catalytically inactive Cas9 (dCas9) or a Cas9 nickase (nCas9), and wherein the first nucleotide sequence of (i) further comprises a nucleotide sequence encoding a nucleobase modifying enzyme fused to the N-terminus of the N-terminal portion of the Cas9 protein.

11. The composition of any one of claims 1-10, wherein the Cas9 protein is a catalytically inactive Cas9 (dCas9) or a Cas9 nickase (nCas9), and wherein the second nucleotide sequence of (ii) further comprises a nucleotide sequence encoding a nucleobase modifying enzyme fused to the C-terminus of the C-terminal portion of the Cas9 protein.

12. The composition of claim 10 or 11, wherein the nucleobase modifying enzyme is a deaminase.

13. The composition of claim 12, wherein the deaminase is a cytosine deaminase.

14. The composition of claim 13, wherein the cytosine deaminase is selected from the group consisting of APOBEC1, APOBEC3, activation induced deaminase (AID), and pmCDA1.

15. The composition of claim 12, wherein the deaminase is an adenosine deaminase.

16. The composition of any one of claims 10-15, wherein the second nucleotide sequence of (ii) further comprises a nucleotide sequence encoding a uracil glycosylase inhibitor (UGI) at the 3′ end of the second nucleotide sequence.

17. The composition of any one of claims 10-15, wherein the first nucleotide sequence of (i) further comprises a nucleotide sequence encoding a uracil glycosylase inhibitor (UGI) at the 5′ end of the first nucleotide sequence.

18. The composition of claim 16 or 17, wherein the UGI comprises the amino acid sequence of any one of SEQ ID NOs: 299-302.

19. The composition of any one of claims 1-18, wherein the first nucleotide sequence and the second nucleotide sequence are on different vectors.

20. The composition of claim 19, wherein each of the different vectors is a genome of a recombinant adeno-associated virus (rAAV).

21. The composition of claim 19 or 20, wherein each of the different vectors is packaged in a separate rAAV particle.

22. A composition, comprising:

(i) a first recombinant adeno associated virus (rAAV) particle comprising a first nucleotide sequence encoding a N-terminal portion of a Cas9 protein fused at its C-terminus to an intein-N, wherein the N-terminal portion of the Cas9 protein comprises a portion of any one of SEQ ID NOs: 1-275 and 394-397 that corresponds to amino acids 1-573 or 1-637 of SEQ ID NO: 1; and

(ii) a second recombinant adeno associated virus (rAAV) particle comprising a second nucleotide sequence encoding an intein-C fused to the N-terminus of a C-terminal portion of

the Cas9 protein, wherein the C-terminal portion of the Cas9 protein comprises a portion of any one of SEQ ID NOs: 1-275 and 394-397 that corresponds to amino acids 574-1368 or 638-1368 of SEQ ID NO: 1; wherein the intein-C comprises the amino acid sequence as set forth in any one of SEQ ID NOs: 353 and 357, and

wherein the first nucleotide sequence or second nucleotide sequence is operably linked to a nucleotide sequence encoding at least one bipartite nuclear localization signal.

23. A cell comprising the composition of any one of claims 1-22.

24. The cell of claim 23, wherein the N-terminal portion of the Cas9 protein and the C-terminal portion of the Cas9 protein are joined together to form the Cas9 protein.

25. The cell of claim 23 or 24, wherein the cell is a prokaryotic cell.

26. The cell of claim 25, wherein the cell is a bacterial cell.

27. The cell of claim 23 or 24, wherein the cell is a eukaryotic cell.

28. The cell of claim 27, wherein the cell is a yeast cell, a plant cell, or a mammalian cell.

29. The cell of claim 27, wherein the cell is a human cell.

30. A kit comprising the composition of any one of claims 1-22.

31. An *in vitro* or *ex vivo* method comprising:

contacting a cell with the composition of any one of claims 1-22, wherein the contacting results in the delivery of the first nucleotide sequence and the second nucleotide sequence into the cell, and wherein the N-terminal portion of the Cas9 protein and the C-terminal portion of the Cas9 protein are joined to form a Cas9 protein.

32. The composition of any one of claims 1-22 for use in therapy, wherein the composition further comprises a nucleotide sequence encoding a gRNA to target the Cas9 protein to a DNA sequence associated with a disease or disorder.

33. Use of the composition of any one of claims 1-22 in the preparation of a medicament for the treatment of a disease or disorder selected from the group consisting of: cystic fibrosis, phenylketonuria, epidermolytic hyperkeratosis (EHK), chronic obstructive pulmonary disease (COPD), Charcot-Marie-Toot disease type 4J, neuroblastoma (NB), von Willebrand disease (vWD), myotonia congenital, hereditary renal amyloidosis, dilated cardiomyopathy, hereditary lymphedema, familial Alzheimer's disease, prion disease, chronic infantile neurologic cutaneous articular syndrome (CINCA), and desmin-related myopathy (DRM), wherein the composition further comprises a nucleotide sequence encoding a gRNA to target the Cas9 protein to a DNA sequence associated with said disease or disorder.