Type VI-E and Type VI-F CRISPR-Cas System and Uses Thereof

BACKGROUND OF THE INVENTION

CRISPR (clustered regularly interspaced short palindromic repeats) is a family of DNA sequences found within the genomes of prokaryotic organisms such as bacteria and archaea. These sequences are understood to be derived from DNA fragments of bacteriophages that have previously infected the prokaryote, and are used to detect and destroy DNA from similar bacterialphages during subsequent infections of the prokaryotes.

CRISPR-associated systems is a set of homologous genes, or Cas genes, some of which encode Cas protein having helicase and nuclease activities. The Cas proteins are enzymes that utilize RNA derived form the CRISPR sequences (crRNA) as guide sequences to recognize and cleave specific strands of polynucleotide (*e.g.*, DNA) that are complementary to the crRNA.

Together, the CRISPR-Cas system constitutes a primitive prokaryotic "immune system" that confers resistance or acquired immunity to foreign pathogenic genetic elements, such as those present within extrachromosomal DNA (*e.g.*, plasmids) and bacterialphages, or foreign RNA encoded by foreign DNA.

In nature, the CRISPR/Cas system appears to be a widespread prokaryotic defense mechanism against foreign genetic materials, and is found in approximately 50% of sequenced bacterial genomes and nearly 90% of sequenced archaea. This prokaryotic system has since been developed to form the basis of a technology known as CRISPR-Cas that found extensive use in numerous eukaryotic organisms including human, in a wide variety of applications including basic biological research, development of biotechnology products, and disease treatment.

The prokaryotic CRISPR-Cas systems comprise an extremely diverse group of proteins effectors, non-coding elements, as well as loci architectures, some examples of which have been engineered and adapted to produce important biotechnologies.

The CRISPR locus structure has been studied in many systems. In these systems, the CRISPR array in the genomic DNA typically comprises an AT-rich leader sequence, followed by short DR sequences separated by unique spacer sequences. These CRISPR DR sequences typically range in size from 28 to 37 bps, though the range can be 23-55 bps. Some DR sequences show dyad symmetry, implying the formation of a secondary structure such as a stem-

loop ("hairpin") in the RNA, while others appear unstructured. The size of spacers in different CRISPR arrays is typically 32-38 bps (with a range of 21-72 bps). There are usually fewer than 50 units of the repeat-spacer sequence in a CRISPR array.

Small clusters of *cas* genes are often found next to such CRISPR repeat-spacer arrays. So far, the 93 identified *cas* genes have been grouped into 35 families, based on sequence similarity of their encoded proteins. Eleven of the 35 families form the so-called *cas* core, which includes the protein families Cas1 through Cas9. A complete CRISPR-Cas locus has at least one gene belonging to the *cas* core.

CRISPR-Cas systems can be broadly divided into two classes - Class 1 systems use a complex of multiple Cas proteins to degrade foreign nucleic acids, while Class 2 systems use a single large Cas protein for the same purpose. The single-subunit effector compositions of the Class 2 systems provide a simpler component set for engineering and application translation, and has thus far been important sources of discovery, engineering, and optimization of novel powerful programmable technologies for genome engineering and beyond.

Class 1 system is further divided into types I, III, and IV; and Class 2 system is divided into types II, V, and VI. These 6 system types are additionally divided into 19 subtypes. Classification is also based on the complement of *cas* genes that are present. Most CRISPR-Cas systems have a Cas1 protein. Many prokaryotes contain multiple CRISPR-Cas systems, suggesting that they are compatible and may share components.

One of the first and best characterized Cas proteins - Cas9 - is a prototypical member of Class 2, type II, and originates from *Streptococcus pyogenes* (SpCas9). Cas9 is a DNA endonuclease activated by a small crRNA molecule that complements a target DNA sequence, and a separate trans-activating CRISPR RNA (tracrRNA). The crRNA consists of a direct repeat (DR) sequence responsible for protein binding to the crRNA and a spacer sequence, which may be engineered to be complementary to any desired nucleic acid target sequence. In this way, CRISPR systems can be programmed to target DNA or RNA targets by modifying the spacer sequence of the crRNA. The crRNA and tracrRNA have been fused to form a single guide RNA (sgRNA) for better practical utility. When combined with Cas9, sgRNA hybridizes with its target DNA, and guides Cas9 to cut the target DNA. Other Cas9 effector protein from other species have also been identified and used similarly, including Cas9 from the *S. thermophilus* CRISPR system. These CRISPR/Cas9 systems have been widely used in numerous eukaryotic

organisms, including baker's yeast (*Saccharomyces cerevisiae*), the opportunistic pathogen *Candida albicans*, zebrafish (*Danio rerio*), fruit flies (*Drosophila melanogaster*), ants (*Harpegnathos saltator* and *Ooceraea biroi*), mosquitoes (*Aedes aegypti*), nematodes (*Caenorhabditis elegans*), plants, mice, monkeys, and human embryos.

Another recently characterized Cas effector protein is Cas12a (formerly known as Cpf1). Cas12a, together with C2c1 and C2c3, are members belonging to Class 2, type V Cas proteins that lack HNH nuclease, but have RuvC nuclease activity. Cas12a which was initially characterized in the CRISPR/Cpf1 system of the bacterium *Francisella novicida*. Its original name reflects the prevalence of its CRISPR-Cas subtype in the *Prevotella* and *Francisella* lineages. Cas12a showed several key differences from Cas9, including: causing a "staggered" cut in double stranded DNA as opposed to the "blunt" cut produced by Cas9, relying on a "T rich" PAM sequence (which provides alternative targeting sites to Cas9) and requiring only a CRISPR RNA (crRNA) and no tracrRNA for successful targeting. Cas12a's small crRNAs are better suited than Cas9 for multiplexed genome editing, as more of them can be packaged in one vector than can Cas9's sgRNAs. Further, the sticky 5' overhangs left by Cas12a can be used for DNA assembly that is much more target-specific than traditional Restriction Enzyme cloning. Finally, Cas12a cleaves DNA 18-23 base pairs downstream from its PAM site, which means no disruption to the nuclease recognition sequence after DNA repair following the creation of double stranded break (DSB) by the NHEJ system, thus Cas12a enables multiple rounds of DNA cleavage, as opposed to the likely one round after Cas9 cleavage because the Cas9 cleavage sequence is only 3 base pairs upstream of the PAM site, and the NHEJ pathway typically results in indel mutations which destroy the recognition sequence, thereby preventing further rounds of cutting. In theory, repeated rounds of DNA cleavage is associated with an increased chance for the desired genomic editing to occur.

More recently, several Class 2, type VI Cas proteins, including Cas13 (also known as C2c2), Cas13b, Cas13c, and Cas13d have been identified, each is an RNA-guided RNase (*i.e.*, these Cas proteins use their crRNA to recognize target RNA sequences, rather than target DNA sequences in Cas9 and Cas12a). Overall, the CRISPR/Cas13 systems can achieve higher RNA digestion efficiency compared to the traditional RNAi and CRISPRi technologies, while simultaneously exhibiting much less off-target cleavage compared to RNAi.

One drawback from these currently identified Cas13 proteins is their relatively large size.

Each of Cas13a, Cas13b, and Cas13c has more than 1100 amino acid residues. Thus it is difficult, if possible at all, to package their coding sequence (about 3.3 kb) and sgRNA, plus any required promoter sequences and translation regulatory sequences, into certain small capacity gene therapy vectors, such as the current most efficient and safest gene therapy vector based on adeno associated virus (AAV), which has a package capacity of about 4.7 kb. Although Cas13d, the smallest Cas13 protein so far, only has about 920 amino acids (*i.e.*, about 2.8 kb coding sequence), and can in theory be packaged into the AAV vector, it has limited use for single-base editing-based gene therapy that depends on using Cas13d-based fusion proteins with single-base editing functions, such as dCas13d-ADAR2DD (which has a coding sequence of about 3.9 kb).

Furthermore, the currently known Cas13 proteins / systems all have non-specific / collateral RNase activity upon activation by crRNA-based target sequence recognition. This activity is particularly strong in Cas13a and Cas13b, and still detectably exists in Cas13d. While this property can be advantageously used in nucleic acid detection methods, the non-specific / collateral RNase activity of these Cas13 proteins constitutes a tremendous potential danger for gene therapy use.

SUMMARY OF THE INVENTION

One aspect of the invention provides a Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-Cas complex, comprising: (1) an RNA guide sequence comprising a spacer sequence capable of hybridizing to a target RNA, and a direct repeat (DR) sequence 3' to the spacer sequence; and, (2) a CRISPR-associated protein (Cas) having an amino acid sequence of any one of SEQ ID NOs: 1-7, or a derivative or functional fragment of said Cas; wherein the Cas, the derivative, and the functional fragment of said Cas, are capable of (i) binding to the RNA guide sequence and (ii) targeting the target RNA, with the proviso that the spacer sequence is not 100% complementary to a naturally-occurring bacterialphage nucleic acid when the complex comprises the Cas of any one of SEQ ID NOs: 1-7 or wherein the target RNA is encoded by a eukaryotic DNA.

In certain embodiments, the DR sequence has substantially the same secondary structure as the secondary structure of any one of SEQ ID NOs: 8-14.

In certain embodiments, the DR sequence is encoded by any one of SEQ ID NOs: 8-14.

In certain embodiments, the target RNA is encoded by a eukaryotic DNA.

In certain embodiments, the eukaryotic DNA is a non-human mammalian DNA, a non-human primate DNA, a human DNA, a plant DNA, an insect DNA, a bird DNA, a reptile DNA, a rodent DNA, a fish DNA, a worm / nematode DNA, a yeast DNA.

In certain embodiments, the target RNA is an mRNA.

In certain embodiments, the spacer sequence is between 15-55 nucleotides, between 25-35 nucleotides, or about 30 nucleotides.

In certain embodiments, the spacer sequence is 90-100% complementary to the target RNA.

In certain embodiments, the derivative comprises conserved amino acid substitutions of one or more residues of any one of SEQ ID NOs: 1-7.

In certain embodiments, the derivative comprises only conserved amino acid substitutions.

In certain embodiments, the derivative has identical sequence to wild-type Cas of any one of SEQ ID NOs: 1-7 in the HEPN domain or the RXXXXH motif.

In certain embodiments, the derivative is capable of binding to the RNA guide sequence hybridized to the target RNA, but has no RNase catalytic activity due to a mutation in the RNase catalytic site of the Cas.

In certain embodiments, the derivative has an N-terminal deletion of no more than 210 residues, and/or a C-terminal deletion of no more than 180 residues.

In certain embodiments, the derivative has an N-terminal deletion of about 180 residues, and/or a C-terminal deletion of about 150 residues.

In certain embodiments, the derivative further comprises an RNA base-editing domain.

In certain embodiments, the RNA base-editing domain is an adenosine deaminase, such as a double-stranded RNA-specific adenosine deaminase (*e.g.*, ADAR1 or ADAR2); apolipoprotein B mRNA editing enzyme; catalytic polypeptide-like (APOBEC); or activation-induced cytidine deaminase (AID).

In certain embodiments, the ADAR has E488Q/T375G double mutation or is ADAR2DD.

In certain embodiments, the base-editing domain is further fused to an RNA-binding domain, such as MS2.

In certain embodiments, the derivative further comprises an RNA methyltransferase, a

RNA demethylase, an RNA splicing modifier, a localization factor, or a translation modification factor.

In certain embodiments, the Cas, the derivative, or the functional fragment comprises a nuclear localization signal (NLS) sequence or a nuclear export signal (NES).

In certain embodiments, targeting of the target RNA results in a modification of the target RNA.

In certain embodiments, the modification of the target RNA is a cleavage of the target RNA.

In certain embodiments, the modification of the target RNA is deamination of an adenosine (A) to an inosine (I).

In certain embodiments, the CRISPR-Cas complex of the invention further comprises a target RNA comprising a sequence capable of hybridizing to the spacer sequence.

Another aspect of the invention provides a fusion protein, comprising (1) the Cas, the derivative thereof, or the functional fragment thereof, of the invention, and (2) a heterologous functional domain.

In certain embodiments, the heterologous functional domain comprises: a nuclear localization signal (NLS), a reporter protein or a detection label (*e.g.*, GST, HRP, CAT, GFP, HcRed, DsRed, CFP, YFP, BFP), a localization signal, a protein targeting moiety, a DNA binding domain (*e.g.*, MBP, Lex A DBD, Gal4 DBD), an epitope tag (*e.g.*, His, myc, V5, FLAG, HA, VSV-G, Trx, *etc*), a transcription activation domain (*e.g.*, VP64 or VPR), a transcription inhibition domain (*e.g.*, KRAB moiety or SID moiety), a nuclease (*e.g.*, FokI), a deamination domain (*e.g.*, ADAR1, ADAR2, APOBEC, AID, or TAD), a methylase, a demethylase, a transcription release factor, an HDAC, a polypeptide having ssRNA cleavage activity, a polypeptide having dsRNA cleavage activity, a polypeptide having ssDNA cleavage activity, a polypeptide having dsDNA cleavage activity, a DNA or RNA ligase, or any combination thereof.

In certain embodiments, the heterologous functional domain is fused N-terminally, C-terminally, or internally in the fusion protein.

Another aspect of the invention provides a conjugate, comprising (1) the Cas, the derivative thereof, or the functional fragment thereof, of the invention, conjugated to (2) a heterologous functional moiety.

In certain embodiments, the heterologous functional moiety comprises: a nuclear localization signal (NLS), a reporter protein or a detection label (*e.g.*, GST, HRP, CAT, GFP, HcRed, DsRed, CFP, YFP, BFP), a localization signal, a protein targeting moiety, a DNA binding domain (*e.g.*, MBP, Lex A DBD, Gal4 DBD), an epitope tag (*e.g.*, His, myc, V5, FLAG, HA, VSV-G, Trx, *etc*), a transcription activation domain (*e.g.*, VP64 or VPR), a transcription inhibition domain (*e.g.*, KRAB moiety or SID moiety), a nuclease (*e.g.*, FokI), a deamination domain (*e.g.*, ADAR1, ADAR2, APOBEC, AID, or TAD), a methylase, a demethylase, a transcription release factor, an HDAC, a polypeptide having ssRNA cleavage activity, a polypeptide having dsRNA cleavage activity, a polypeptide having ssDNA cleavage activity, a polypeptide having dsDNA cleavage activity, a DNA or RNA ligase, or any combination thereof.

In certain embodiments, the heterologous functional moiety is conjugated N-terminally, C-terminally, or internally with respect to the Cas, the derivative thereof, or the functional fragment thereof.

Another aspect of the invention provides a polynucleotide encoding any one of SEQ ID NOs: 1-7, or a derivative thereof, or a functional fragment thereof, or a fusion protein thereof, provided that the polynucleotide is not any one of SEQ ID NOs: 15-21.

In certain embodiments, the polynucleotide is codon-optimized for expression in a cell.

In certain embodiments, the cell is a eukaryotic cell.

Another aspect of the invention provides a non-naturally occurring polynucleotide comprising a derivative of any one of SEQ ID NOs: 8-14, wherein said derivative (i) has one or more (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10) nucleotides additions, deletions, or substitutions compared to any one of SEQ ID NOs: 8-14; (ii) has at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 97% sequence identity to any one of SEQ ID NOs: 8-14; (iii) hybridize under stringent conditions with any one of SEQ ID NOs: 8-14 or any of (i) and (ii); or (iv) is a complement of any of (i) - (iii), provided that the derivative is not any one of SEQ ID NOs: 8-14, and that the derivative encodes an RNA (or is an RNA) that has maintained substantially the same secondary structure as any of the RNA encoded by SEQ ID NOs: 8-14.

In certain embodiments, the derivative functions as a DR sequence for any one of the Cas, the derivative thereof, or the functional fragment thereof, of the invention.

Another aspect of the invention provides a vector comprising the polynucleotide of the invention.

In certain embodiments, the polynucleotide is operably linked to a promoter and optionally an enhancer.

In certain embodiments, the promoter is a constitutive promoter, an inducible promoter, a ubiquitous promoter, or a tissue specific promoter.

In certain embodiments, the vector is a plasmid.

In certain embodiments, the vector is a retroviral vector, a phage vector, an adenoviral vector, a herpes simplex viral (HSV) vector, an AAV vector, or a lentiviral vector.

In certain embodiments, the AAV vector is a recombinant AAV vector of the serotype AAV1, AAV2, AAV4, AAV5, AAV6, AAV7, AAVrh74, AAV8, AAV9, AAV10, AAV 11, AAV 12, or AAV 13.

Another aspect of the invention provides a delivery system comprising (1) a delivery vehicle, and (2) the CRISPR-Cas complex of the invention, the fusion protein of the invention, the conjugate of the invention, the polynucleotide of the invention, or the vector of the invention.

In certain embodiments, the delivery vehicle is a nanoparticle, a liposome, an exosome, a microvesicle, or a gene-gun.

Another aspect of the invention provides a cell or a progeny thereof, comprising the CRISPR-Cas complex of the invention, the fusion protein of the invention, the conjugate of the invention, the polynucleotide of the invention, or the vector of the invention.

In certain embodiments, the cell or progeny thereof is a eukaryotic cell (*e.g.*, a non-human mammalian cell, a human cell, or a plant cell) or a prokaryotic cell (*e.g.*, a bacteria cell).

Another aspect of the invention provides a non-human multicellular eukaryote comprising the cell of the invention.

In certain embodiments, the non-human multicellular eukaryote is an animal (*e.g.*, rodent or primate) model for a human genetic disorder.

Another aspect of the invention provides a method of modifying a target RNA, the method comprising contacting the target RNA with the CRISPR-Cas complex of the invention, wherein the spacer sequence is complementary to at least 15 nucleotides of the target RNA; wherein the Cas, the derivative, or the functional fragment associates with the RNA guide sequence to form the complex; wherein the complex binds to the target RNA; and wherein upon binding of the complex to the target RNA, the Cas, the derivative, or the functional fragment

modifies the target RNA.

In certain embodiments, the target RNA is modified by cleavage by the Cas.

In certain embodiments, the target RNA is modified by deamination by a derivative comprising a Double-stranded RNA-specific adenosine deaminase.

In certain embodiments, the target RNA is an mRNA, a tRNA, an rRNA, a non-coding RNA, an lncRNA, or a nuclear RNA.

In certain embodiments, upon binding of the complex to the target RNA, the Cas, the derivative, and the functional fragment does not exhibit substantial (or detectable) collateral RNase activity.

In certain embodiments, the target RNA is within a cell.

In certain embodiments, the cell is a cancer cell.

In certain embodiments, the cell is infected with an infectious agent.

In certain embodiments, the infectious agent is a virus, a prion, a protozoan, a fungus, or a parasite.

In certain embodiments, the CRISPR-Cas complex is encoded by a first polynucleotide encoding any one of SEQ ID NOs: 1-7, or a derivative or functional fragment thereof, and a second polynucleotide comprising any one of SEQ ID NOs: 8-14 and a sequence encoding a spacer RNA capable of binding to the target RNA, wherein the first and the second polynucleotides are introduced into the cell.

In certain embodiments, the first and the second polynucleotides are introduced into the cell by the same vector.

In certain embodiments, the method causes one or more of: (i) *in vitro* or *in vivo* induction of cellular senescence; (ii) *in vitro* or *in vivo* cell cycle arrest; (iii) *in vitro* or *in vivo* cell growth inhibition and/or cell growth inhibition; (iv) *in vitro* or *in vitro* induction of anergy; (v) *in vitro* or *in vitro* induction of apoptosis; and (vi) *in vitro* or *in vitro* induction of necrosis.

Another aspect of the invention provides a method of treating a condition or disease in a subject in need thereof, the method comprising administering to the subject a composition comprising the CRISPR-Cas complex of the invention or a polynucleotide encoding the same; wherein the spacer sequence is complementary to at least 15 nucleotides of a target RNA

associated with the condition or disease; wherein the Cas, the derivative, or the functional fragment associates with the RNA guide sequence to form the complex; wherein the complex binds to the target RNA; and wherein upon binding of the complex to the target RNA, the Cas, the derivative or the functional fragment cleaves the target RNA, thereby treating the condition or disease in the subject.

In certain embodiments, the condition or disease is a cancer or an infectious disease.

In certain embodiments, the cancer is Wilms' tumor, Ewing sarcoma, a neuroendocrine tumor, a glioblastoma, a neuroblastoma, a melanoma, skin cancer, breast cancer, colon cancer, rectal cancer, prostate cancer, liver cancer, renal cancer, pancreatic cancer, lung cancer, biliary cancer, cervical cancer, endometrial cancer, esophageal cancer, gastric cancer, head and neck cancer, medullary thyroid carcinoma, ovarian cancer, glioma, lymphoma, leukemia, myeloma, acute lymphoblastic leukemia, acute myelogenous leukemia, chronic lymphocytic leukemia, chronic myelogenous leukemia, Hodgkin's lymphoma, non-Hodgkin's lymphoma, or urinary bladder cancer.

In certain embodiments, the method is an *in vitro* method, an *in vivo* method, or an ex vivo method.

Another aspect of the invention provides a cell or a progeny thereof, obtained by the method of the invention, wherein the cell and the progeny comprises a non-naturally existing modification (*e.g.*, a non-naturally existing modification in a transcribed RNA of the cell / progeny).

Another aspect of the invention provides a method to detect the presence of a target RNA, the method comprising contacting the target RNA with a composition comprising a fusion protein of the invention, or a conjugate of the invention, or a polynucleotide encoding the fusion protein, wherein the fusion protein or the conjugate comprises a detectable label (*e.g.*, one that can be detected by fluorescence, Northern blot, or FISH) and a complexed spacer sequence capable of binding to the target RNA.

Another aspect of the invention provides a eukaryotic cell comprising a Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-Cas complex, said CRISPR-Cas complex comprising: (1) an RNA guide sequence comprising a spacer sequence capable of hybridizing to a target RNA, and a direct repeat (DR) sequence 3' to the spacer sequence; and,

(2) a CRISPR-associated protein (Cas) having an amino acid sequence of any one of SEQ ID NOs: 1-7, or a derivative or functional fragment of said Cas; wherein the Cas, the derivative, and the functional fragment of said Cas, are capable of (i) binding to the RNA guide sequence and (ii) targeting the target RNA.

It should be understood that any one embodiment of the invention described herein, including those described only in the examples or claims, or only in one aspects / sections below, can be combined with any other one or more embodiments of the invention, unless explicitly disclaimed or improper.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a schematic (not to scale) illustration of the genomic loci of the representative Cas13e and Cas13f families members. The Cas coding sequences (long bars with pointed end), followed by the multiple nearby direct repeat (DR) (short bars) and spacer sequences (diamonds) are shown.

FIG. 2 shows putative secondary structures of the DR sequences associated with the respective Cas13e and Cas13f proteins. Their coding sequences, from left to right, are represented by SEQ ID NOs: 8-14, respectively.

FIG. 3 shows a phylogenetic tree for the newly discovered Cas13e and Cas13f effector proteins of the invention, as well as the related previously discovered Cas13a, Cas13b, Cas13c, and Cas13d effector proteins.

FIG. 4 shows the domain structures for the Cas13a-Cas13f proteins. The overall sizes, and the locations of the two RXXXXH motifs on each representative member of the Cas proteins are indicated.

FIG. 5 shows a predicted 3D structure of the Cas13e.1 effector protein.

FIG. 6 is a schematic drawing showing that the three plasmids, encoding (1) a Cas13e effector protein, (2) a coding sequence for the guide RNA (gRNA) which can produce the guide RNA that is complementary to the mCherry mRNA and that can form a complex with the Cas13e effector protein, and (3) the mCherry reporter gene, respectively, can be transfected to a cell to express their respective gene products, resulting in the degradation of the reporter mCherry mRNA.

FIG. 7 shows knock-down of mCherry mRNA by guide RNA complementary to the mCherry mRNA, as evidenced by reduced mCherry expression under fluorescent microscope. As a negative control, a non-targeting (NT) guide RNA that does not hybridize with / bind to the mCherry mRNA failed to knock-down mCherry expression.

FIG. 8 shows about 75% knock-down of mCherry expression in experiments in FIG. 6.

FIG. 9 shows that Cas13e utilizes a guide RNA having a DR sequence at the 3' end (as opposed to a DR sequence at the 5'-end of the guide RNA).

FIG. 10 shows the correlation between spacer sequence length and specific (guide RNA-dependent) RNase activity against target RNA relative to non-targeting (NT) control.

FIG. 11 shows the correlation between spacer sequence length and non-specific / collateral (guide RNA-independent) RNase activity against target RNAs relative to non-targeting (NT) control.

FIG. 12 shows that dCas13e.1-ADAR2DD fusion has RNA base editing activity. Specifically, three plasmids, encoding (1) a dCas13e (RNase dead) protein fused to the single-base RNA editor ADAR2DD, (2) a coding sequence for the guide RNA (gRNA) which can produce the guide RNA that is complementary to a mutant mCherry mRNA having a G-to-A point mutation and that can form a complex with the dCas13e effector protein, and (3) the mutant mCherry reporter gene encoding the mCherry mRNA having the G-to-A point mutation, respectively, can be transfected to a cell to express their respective gene products. The mutant mCherry mRNA normally cannot produced a fluorescent mCherry protein due to the point mutation. Upon guide RNA binding to the mutant mCherry mRNA, the fused ADAR2DD base editor converts A to I (G equivalent), thus restoring the ability of the mRNA to encode a fluorescent mCherry protein.

FIG. 13 shows restored expression of mCherry as a result of successful RNA base editing. In the Experiment in FIG. 12, plasmid encoding mutant mCherry (mCherry*) alone failed to express fluorescent mCherry. Plasmid encoding dCas13e-ADAR2DD base editor alone also failed to express fluorescent mCherry. Plasmid encoding either gRNA-1 or gRNA-2 alone (which also expresses a GFP reporter) also failed to express fluorescent mCherry, though GFP was expressed prominently. However, when all three plasmids were transfected into the same cell, significant fluorescent mCherry expression was observed (together with GFP reporter

expression).

FIG. 14 shows the relevant segment of the mutant mCherry gene having the premature stop codon TAG, the sequence for the two gRNA that can be complexed with the dCas13e-ADAR2DD RNA base editor, and the "corrected" TGG codon.

FIG. 15 is a schematic (not to scale) drawing showing the series of progressive C-terminal deletion constructs for dCas13e.1 fused to the ADAR2DD RNA base editor (shown as "ADAR2"), as well as other transcriptional control elements.

FIG. 16 shows the percentage results of mCherry mutant conversion back to wild-type mCherry, for the series of C-terminal deletion mutants in FIG. 15.

FIG. 17 is a schematic (not to scale) drawing showing the series of progressive C-terminal and optional N-terminal deletion constructs for dCas13e.1 fused to the ADAR2DD RNA base editor.

FIG. 18 shows the percentage results of mCherry mutant conversion back to wild-type mCherry, for selected C- and N-terminal deletion mutants in FIG. 17.

FIG. 19 shows the series of plasmids encoding Cas13a, Cas13b, Cas13d, Cas13e.1 and Cas13f.1, the mCherry reporter gene, as well as either the ANXA4-targeting gRNA coding sequence, or a non-targeting gRNA as control.

FIG. 20 shows efficient knock-down of ANXA4 expression by Cas13e.1, Cas13f.1, Cas13a, as well as Cas13d.

DETAILED DESCRIPTION OF THE INVENTION

1. Overview

The invention described herein provides novel Class 2, type VI Cas effector proteins, sometimes referred herein as Cas13e and Cas13f. The novel Cas13 proteins of the invention are much smaller than the previously discovered Cas13 effector proteins (Cas13a-Cas13d), such that they can be easily packaged with their crRNA coding sequences into small capacity gene therapy vectors, such as the AAV vectors. Further, the newly discovered Cas13e and Cas13f effector proteins are more potent in knocking down RNA target sequences, and more efficient in RNA single base editing, as compared to the Cas13a, Cas13b, and Cas13d effector proteins, while

exhibiting negligible non-specific / collateral RNase activity upon activation by crRNA-based target recognition, except when the spacer sequence is within a specific narrow range (*e.g.*, about 30 nucleotide). Thus these new Cas proteins are ideally suited for gene therapy.

Thus in the first aspect, the invention provides Cas13e and Cas13f effector proteins, such as those with amino acid sequences of SEQ ID NOs: 1-7, or orthologs, homologs, the various derivatives (described herein below), functional fragments thereof (described herein bellow), wherein said orthologs, homologs, derivatives and functional fragments have maintained at least one function of any one of the proteins of SEQ ID NOs: 1-7. Such functions include, but are not limited to, the ability to bind a guide RNA / crRNA of the invention (described herein below) to form a complex, the RNase activity, and the ability to bind to and cleave a target RNA at a specific site, under the guidance of the crRNA that is at least partially complementary to the target RNA.

In certain embodiments, the Cas13e or Cas13f effector proteins of the invention can be: (i) any one of SEQ ID NOs: 1-7; (ii) a derivative having one or more amino acids (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 residues) of addition, deletion, and/or substitution (*e.g.*, conserved substitution) of any one of SEQ ID NOs: 1-7; or (iii) a derivative having amino acid sequence identity of at least about 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% compared to any one of SEQ ID NOs: 1-7.

In certain embodiments, the Cas13e and Cas13f effector proteins, orthologs, homologs, derivatives and functional fragments thereof are not naturally existing, *e.g.*, having at least one amino acid difference compared to a naturally existing sequence.

In a related aspect, the invention provides additional derivatives Cas13e and Cas13f effector proteins based on any one of SEQ ID NOs: 1-7, or the above orthologs, homologs, derivatives and functional fragments thereof, which comprises another covalently or non-covalently linked protein or polypeptide or other molecules (such as detection reagents or drug / chemical moieties). Such other proteins / polypeptides / other molecules can be linked through, for example, chemical coupling, gene fusion, or other non-covalent linkage (such as biotin-streptavidin binding). Such derived proteins do not affect the function of the original protein, such as the ability to bind a guide RNA / crRNA of the invention (described herein below) to form a complex, the RNase activity, and the ability to bind to and cleave a target RNA at a specific site, under the guidance of the crRNA that is at least partially complementary to the

target RNA.

Such derivation may be used, for example, to add a nuclear localization signal (NLS, such as SV40 large T antigen NLS) to enhance the ability of the subject Cas13e and Cas13f effector proteins to enter cell nucleus. Such derivation can also be used to add a targeting molecule or moiety to direct the subject Cas13e and Cas13f effector proteins to specific cellular or subcellular locations. Such derivation can also be used to add a detectable label to facilitate the detection, monitoring, or purification of the subject Cas13e and Cas13f effector proteins. Such derivation can further be used to add a deamination enzyme moiety (such as one with adenine or cytosine deamination activity) to facilitate RNA base editing.

The derivation can be through adding any of the additional moieties at the N- or C-terminal of the subject Cas13e and Cas13f effector proteins, or internally (*e.g.*, internal fusion or linkage through side chains of internal amino acids).

In a related second aspect, the invention provides conjugates of the subject Cas13e and Cas13f effector proteins based on any one of SEQ ID NOs: 1-7, or the above orthologs, homologs, derivatives and functional fragments thereof, which are conjugated with moieties such as other proteins or polypeptides, detectable labels, or combinations thereof. Such conjugated moieties may include, without limitation, localization signals, reporter genes (*e.g.*, GST, HRP, CAT, GFP, HcRed, DsRed, CFP, YFP, BFP), labels (*e.g.*, fluorescent dye such as FITC, or DAPI), NLS, targeting moieties, DNA binding domains (*e.g.*, MBP, Lex A DBD, Gal4 DBD), epitope tags (*e.g.*, His, myc, V5, FLAG, HA, VSV-G, Trx, *etc*), transcription activation domains (*e.g.*, VP64 or VPR), transcription inhibition domains (*e.g.*, KRAB moiety or SID moiety), nucleases (*e.g.*, FokI), deamination domain (*e.g.*, ADAR1, ADAR2, APOBEC, AID, or TAD), methylase, demethylase, transcription release factor, HDAC, ssRNA cleavage activity, dsRNA cleavage activity, ssDNA cleavage activity, dsDNA cleavage activity, DNA or RNA ligase, any combination thereof, *etc*.

For example, the conjugate may include one or more NLSs, which can be located at or near N-terminal, C-terminal, internally, or combination thereof. The linkage can be through amino acids (such as D or E, or S or T), amino acid derivatives (such as Ahx, β-Ala, GABA or Ava), or PEG linkage.

In certain embodiments, conjugations do not affect the function of the original protein,

such as the ability to bind a guide RNA / crRNA of the invention (described herein below) to form a complex, the RNase activity, and the ability to bind to and cleave a target RNA at a specific site, under the guidance of the crRNA that is at least partially complementary to the target RNA.

In a related third aspect, the invention provides fusions of the subject Cas13e and Cas13f effector proteins based on any one of SEQ ID NOs: 1-7, or the above orthologs, homologs, derivatives and functional fragments thereof, which fusions are with moieties such as localization signals, reporter genes (*e.g.*, GST, HRP, CAT, GFP, HcRed, DsRed, CFP, YFP, BFP), NLS, protein targeting moieties, DNA binding domains (*e.g.*, MBP, Lex A DBD, Gal4 DBD), epitope tags (*e.g.*, His, myc, V5, FLAG, HA, VSV-G, Trx, *etc*), transcription activation domains (*e.g.*, VP64 or VPR), transcription inhibition domains (*e.g.*, KRAB moiety or SID moiety), nucleases (*e.g.*, FokI), deamination domain (*e.g.*, ADAR1, ADAR2, APOBEC, AID, or TAD), methylase, demethylase, transcription release factor, HDAC, ssRNA cleavage activity, dsRNA cleavage activity, ssDNA cleavage activity, dsDNA cleavage activity, DNA or RNA ligase, any combination thereof, *etc*.

For example, the fusion may include one or more NLSs, which can be located at or near N-terminal, C-terminal, internally, or combination thereof. In certain embodiments, conjugations do not affect the function of the original protein, such as the ability to bind a guide RNA / crRNA of the invention (described herein below) to form a complex, the RNase activity, and the ability to bind to and cleave a target RNA at a specific site, under the guidance of the crRNA that is at least partially complementary to the target RNA.

In a fourth aspect, the invention provides an isolated polynucleotide, comprising: (i) any one of SEQ ID NOs: 8-14; (ii) a polynucleotide having 1, 2, 3, 4, or 5 nucleotides of deletion, addition, and/or substitution compared to any one of SEQ ID NOs: 8-14; (iii) a polynucleotide sharing at least 80%, 85%, 90%, 95% sequence identity with any one of SEQ ID NOs: 8-14; (iv) a polynucleotide that hybridize under stringent condition with any one of the polynucleotide of (i) - (iii) or a complement thereof; (v) a complement sequence of any polynucleotide of (i) - (iii).

Any polynucleotide of (ii) - (iv) has maintained the function of the original SEQ ID NOs: 8-14, which is to encode a direct repeat (DR) sequence of a crRNA in the subject Cas13e or Cas13f system.

As used herein, "direct repeat sequence" may refer to the DNA coding sequence in the CRISPR locus, or to the RNA encoded by the same in crRNA. Thus when any of SEQ ID NOs: 8-14 is referred to in the context of an RNA molecule, such as crRNA, each T is understood to represent a U.

Thus in certain embodiments, the isolated polynucleotide is a DNA, which encodes a DR sequence for a crRNA of the subject Cas13e and Cas13f system.

In certain other embodiments, the isolated polynucleotide is an RNA, which is a DR sequence for a crRNA of the subject Cas13e and Cas13f system.

In a fifth aspect, the invention provides a complex comprising: (i) a protein composition that can be any one of the subject Cas13e or Cas13f effector protein, or orthologs, homologs, derivatives, conjugates, functional fragments thereof, conjugates thereof, or fusions thereof; and (ii) a polynucleotide composition, comprising an isolated polynucleotide described in the 4th aspect of the invention (*e.g.*, a DR sequence), and a spacer sequence complementary to at least a portion of a target RNA. In certain embodiments, the DR sequence is at the 3' end of the spacer sequence.

In some embodiments, the polynucleotide composition is the guide RNA / crRNA of the subject Cas13e or Cas13f system, which does not include a tracrRNA.

In certain embodiments, for use with Cas13e and Cas13f effector proteins, homologs, orthologs, derivatives, fusions, conjugates, or functional fragments thereof having RNase activity, the spacer sequence is at least about 10 nucleotides, or between 10-60, 15-50, 20-50, 25-40, 25-50, or 19-50 nucleotides. In certain embodiments, for use with Cas13e and Cas13f effector proteins, homologs, orthologs, derivatives, fusion, conjugates, or functional fragments thereof having no RNase activity but ability to bind guide RNA and a target RNA complementary to the guide RNA, the spacer sequence is at least about 10 nucleotides, or between about 10-200, 15-180, 20-150, 25-125, 30-110, 35-100, 40-80, 45-60, 50-55, or about 50 nucleotides.

In certain embodiments, the DR sequence is between 15-36, 20-36, 22-36, or about 36 nucleotides. In certain embodiments, the DR sequence in the guide RNA has substantially the same secondary structure (including stems, bulges, and loop) as the RNA version of any one of SEQ ID NOs: 8-14.

In certain embodiments, the guide RNA is about 36 nucleotides longer than any of the

spacer sequence lengths above, such as between 45-96, 55-86, 60-86, 62-86, or 63-86 nucleotides.

In a sixth aspect, the invention provides an isolated polynucleotide comprising: (i) a polynucleotide encoding any one of the Cas13e or Cas13f effector proteins of SEQ ID NOs: 1-7, or orthologs, homologs, derivatives, functional fragments, fusions thereof; (ii) a polynucleotide of any one of SEQ ID NOs: 8-14; or (iii) a polynucleotide comprising (i) and (ii).

In some embodiments, the polynucleotide is not naturally occurring / naturally existing, such as excluding SEQ ID NOs: 15-21.

In some embodiments, the polynucleotide is codon-optimized for expression in a prokaryote. In some embodiments, the polynucleotide is codon-optimized for expression in a eukaryote, such as in human or human cell.

In a seventh aspect, the invention provides a vector comprising or encompassing any of the polynucleotide of the sixth aspect. The vector can be a cloning vector, or an expression vector. The vector can be a plasmid, phagemid, or cosmid, just to name a few. In certain embodiments, the vector can be used to express the polynucleotide in a mammalian cell, such as a human cell, any one of the Cas13e or Cas13f effector proteins of SEQ ID NOs: 1-7, or orthologs, homologs, derivatives, functional fragments, fusions thereof; or any of the polynucleotide of the 4th aspect; or any of the complex of the 5th aspect.

In an eighth aspect, the invention provides a host cell comprising any of the polynucleotide of the 4th or 6th aspect, and/or the vector of the 7th aspect of the invention. The host cell can be a prokaryote such as *E. coli*, or a cell from a eukaryote such as yeast, insect, plant, animal (*e.g.*, mammal including human and mouse). The host cell can be isolated primary cell (such as bone marrow cells for ex vivo therapy), or established cell lines such as tumor cell lines, 293T cells, or stem cells, iPCs, *etc.*

In a related aspect, the invention provides a eukaryotic cell comprising a Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-Cas complex, said CRISPR-Cas complex comprising: (1) an RNA guide sequence comprising a spacer sequence capable of hybridizing to a target RNA, and a direct repeat (DR) sequence 3' to the spacer sequence; and, (2) a CRISPR-associated protein (Cas) having an amino acid sequence of any one of SEQ ID NOs: 1-7, or a derivative or functional fragment of said Cas; wherein the Cas, the derivative, and

the functional fragment of said Cas, are capable of (i) binding to the RNA guide sequence and (ii) targeting the target RNA.

In a ninth aspect, the invention provides a composition comprising: (i) a first (protein) composition selected from any one of the Cas13e or Cas13f effector proteins of SEQ ID NOs: 1-7, or orthologs, homologs, derivatives, conjugates, functional fragments, fusions thereof; and (ii) a second (nucleotide) composition comprising an RNA encompassing a guide RNA / crRNA, particularly a spacer sequence, or a coding sequence for the same. The guide RNA may comprise a DR sequence, and a spacer sequence which can complement or hybridize with a target RNA. The guide RNA can form a complex with the first (protein) composition of (i). In some embodiment, the DR sequence can be the polynucleotide of the 4th aspect of the invention. In some embodiment, the DR sequence can be at the 3'-end of the guide RNA. In some embodiments, the composition (such as (i) and/or (ii)) is non-naturally occurring or modified from a naturally occurring composition. In some embodiments, at least a component of the composition is non-naturally occurring or modified from a naturally occurring component of the composition. In some embodiments, the target sequence is an RNA from a prokaryote or a eukaryote, such as a non-naturally existing RNA. The target RNA may be present inside a cell, such as in the cytosol or inside an organelle. In some embodiments, the protein composition may have an NLS that can be located at its N- or C-terminal, or internally.

In a tenth aspect, the invention provides a composition comprising one or more vectors of the 7th aspect of the invention, said one or more vectors comprise: (i) a first polynucleotide that encodes any one of the Cas13e or Cas13f effector proteins of SEQ ID NOs: 1-7, or orthologs, homologs, derivatives, functional fragments, fusions thereof; optionally operably linked to a first regulatory element; and (ii) a second polynucleotide that encodes a guide RNA of the invention; optionally operably linked to a second regulatory element. The first and the second polynucleotides can be on different vectors, or on the same vector. The guide RNA can form a complex with the protein product encoded by the first polynucleotide, and comprises a DR sequence (such as any one of the 4th aspect) and a spacer sequence that can bind to / complement with a target RNA. In some embodiments, the first regulatory element is a promoter, such as an inducible promoter. In some embodiments, the second regulatory element is a promoter, such as an inducible promoter. In some embodiments, the composition (such as (i) and/or (ii)) is non-naturally occurring or modified from a naturally occurring composition. In some embodiments,

at least a component of the composition is non-naturally occurring or modified from a naturally occurring component of the composition. In some embodiments, the target sequence is an RNA from a prokaryote or a eukaryote, such as a non-naturally existing RNA. The target RNA may be present inside a cell, such as in the cytosol or inside an organelle. In some embodiments, the protein composition may have an NLS that can be located at its N- or C-terminal, or internally.

In some embodiments, the vector is a plasmid. In some embodiment, the vector is a viral vector based on a retrovirus, a replication incompetent retrovirus, adenovirus, replication incompetent adenovirus, or AAV. In some embodiments, the vector can self-replicate in a host cell (*e.g.*, having a bacterial replication origin sequence). In some embodiments, the vector can integrate into a host genome and be replicated therewith. In some embodiment, the vector is a cloning vector. In some embodiment, the vector is an expression vector.

The invention further provides a delivery composition for delivering any of the Cas13e or Cas13f effector proteins of SEQ ID NOs: 1-7, or orthologs, homologs, derivatives, conjugates, functional fragments, fusions thereof of the 1st-3rd aspects of the invention; the polynucleotide of the 4th and/or 6th aspect of the invention; the complex of the 5th aspect of the invention; the vector of the 7th aspect of the invention; the cell of the 8th aspect of the invention, and the composition of the 9th and/or 10th aspects of the invention. The delivery can be through any one known in the art, such as transfection, lipofection, electroporation, gene gun, microinjection, sonication, calcium phosphate transfection, cation transfection, viral vector delivery, *etc.*, using vehicles such as liposome(s), nanoparticle(s), exosome(s), microvesicle(s), a gene-gun or one or more viral vector(s).

The invention further provides a kit comprising any one or more of the following: any of the Cas13e or Cas13f effector proteins of SEQ ID NOs: 1-7, or orthologs, homologs, derivatives, conjugates, functional fragments, fusions thereof of the 1st-3rd aspects of the invention; the polynucleotide of the 4th and/or 6th aspect of the invention; the complex of the 5th aspect of the invention; the vector of the 7th aspect of the invention; the cell of the 8th aspect of the invention, and the composition of the 9th and/or 10th aspects of the invention. In some embodiments, the kit may further comprise an instruction for how to use the kit components, and/or how to obtain additional components from 3rd party for use with the kit components. Any component of the kit can be stored in any suitable container.

With the inventions generally described herein above, more detailed descriptions for the

various aspects of the invention are provided in separate sections below. However, it should be understood that, for simplicity and to reduce redundancy, certain embodiments of the invention are only described under one section or only described in the claims or examples. Thus it should also be understood that any one embodiment of the invention, including those described only under one aspect, section, or only in the claims or examples, can be combined with any other embodiment of the invention, unless specifically disclaimed or the combination is improper.

2. Novel Class 2, Type VI CRISPR RNA-Guided RNases, and Derivatives Thereof

In one aspect, the invention described herein provides two novel families of CRISPR Class 2, type VI effectors having two strictly conserved RX4-6H (RXXXXH) motifs, characteristic of Higher Eukaryotes and Prokaryotes Nucleotide-binding (HEPN) domains. Similar CRISPR Class 2, type VI effectors that contain two HEPN domains have been previously characterized and include, for example, CRISPR Cas13a (C2c2), Cas13b, Cas13c, and Cas13d.

HEPN domains have been shown to be RNase domains and confer the ability to bind to and cleave target RNA molecule. The target RNA may be any suitable form of RNA, including but not limited to mRNA, tRNA, ribosomal RNA, non-coding RNA, lncRNA (long non-coding RNA), and nuclear RNA. For example, in some embodiments, the Cas proteins recognize and cleave RNA targets located on the coding strand of open reading frames (ORFs).

In one embodiment, the disclosure provides two families of CRISPR Class 2, type VI effectors, referred to herein generally as Type VI-E and VI-F CRISPR-Cas effector proteins, Cas13e or Cas13f. Direct comparison of the Type VI-E and VI-F CRISPR-Cas effector proteins with the effector of these other systems shows that Type VI-E and VI-F CRISPR-Cas effector proteins are significantly smaller (*e.g.*, about 20% fewer amino acids) than even the smallest previously identified Type VI-D / Cas13d effectors (see FIG. 4), and have less than 30% sequence similarity in one to one sequence alignments to other previously described effector proteins, including the phylogenetically closest relatives Cas13b (see FIG. 3).

These two newly-identified families of CRISPR Class 2, type VI effectors can be used in a variety of applications, and are particularly suitable for therapeutic applications since they are significantly smaller than other effectors (*e.g.*, CRISPR Cas13a, Cas13b, Cas13c, and Cas13d

effectors) which allows for the packaging of the nucleic acids encoding the effectors and their guide RNA coding sequences into delivery systems having size limitations, such as the AAV vectors. Further, the lack of detectable collateral / non-specific RNase activity at selected range of spacer sequence lengths (such as about 30 nucleotides, see FIG. 11), upon activation of the specific RNase activity, makes these Cas effectors less prong to (if not immune from) potentially dangerous generalized off-target RNA digestion in target cells that are desirably not destroyed. On the other hand, at other selected spacer lengths such as about 30 nucleotides, significant collateral RNase activity exists for these Cas effectors, thus the subject Cas effectors can also be used in utilities depending on such collateral RNase activity.

In bacteria, the Type VI-E and VI-F CRISPR-Cas systems include a single effector (approximately 775 residues and 790 residues, respectively) within close proximity to a CRISPR array (see FIG. 1). The CRISPR array includes direct repeat (DR) sequences typically 36 nucleotides in length, which are generally well conserved, both in sequences and secondary structures (see FIG. 2).

Data provided herein demonstrated that the crRNA is processed from the 5'-end, such that the DR sequences end up at the 3'-end of the mature crRNA.

The spacers contained in the Cas13e and Cas13f CRISPR arrays are most commonly 30 nucleotides in length, with the majority of variation in length contained in the range of 29 to 30 nucleotides. However, a wide range of spacer length may be tolerated. For example, for use in a functional Cas13e or Cas13f effector protein, or homologs, orthologs, derivatives, fusions, conjugates, or functional fragment thereof, the spacer can be between 10-60 nucleotides, 20-50 nucleotides, 25-45 nucleotides, 25-35 nucleotides, or about 27, 28, 29, 30, 31, 32, or 33 nucleotides. For use in dCas version of any of the above, however, the spacer can be between 10-200 nucleotides, 20-150 nucleotides, 25-100 nucleotides, 25-85 nucleotides, 35-75 nucleotides, 45-60 nucleotides, or about 46, 47, 48, 49, 50, 51, 52, 53, 54, or 55 nucleotides.

Exemplary Type VI-E and VI-F CRISPR-Cas effector proteins are provided in the table below.

Cas13e.1	MAQVSKQTSKKRELSIDEYQGARKWCFTIAFNKALVNRDKNDGLFVESLLRHEKYSKHDWY DEDTRALIKCSTQAANAKAEALRNYFSHYRHSPGCLTFTAEDELRTIMERAYERAIFECRR RETEVIIEFPSLFEGDRITTAGVVFFVSFFVERRVLDRLYGAVSGLKKNEGQYKLTRKALS MYCLKDSRFTKAWDKRVLLFRDILAQLGRIPAEAYEYYHGEQGDKKRANDNEGTNPKRHKD KFIEFALHYLEAQHSEICFGRRHIVREEAGAGDEHKKHRTKGKVVVDFSKKDEDQSYYISK

	NNVIVRIDKNAGPRSYRMGLNELKYLVLLSLQGKGDDAIAKLYRYRQHVENILDVVKVTDK DNHVFLPRFVLEQHGIGRKAFKQRIDGRVKHVRGVWEKKKAATNEMTLHEKARDILQYVNE NCTRSFNPGEYNRLLVCLVGKDVENFQAGLKRLQLAERIDGRVYSIFAQTSTINEMHQVVC DQILNRLCRIGDQKLYDYVGLGKKDEIDYKQKVAWFKEHISIRRGFLRKKFWYDSKKGFAK LVEEHLESGGGQRDVGLDKKYYHIDAIGRFEGANPALYETLARDRLCLMMAQYFLGSVRKE LGNKIVWSNDSIELPVEGSVGNEKSIVFSVSDYGKLYVLDDAEFLGRICEYFMPHEKGKIR YHTVYEKGFRAYNDLQKKCVEAVLAFEEKVVKAKKMSEKEGAHYIDFREILAQTMCKEAEK TAVNKVRRAFFHHHLKFVIDEFGLFSDVMKKYGIEKEWKFPVK* (SEQ ID NO: 1)
Cas13e.2	MKVENIKEKSKKAMYLINHYEGPKKWCFAIVLNRACDNYEDNPHLFSKSLLEFEKTSRKDW FDEETRELVEQADTEIQPNPNLKPNTTANRKLKDIRNYFSHHYHKNECLYFKNDDPIRCIM EAAYEKSKIYIKGKQIEQSDIPLPELFESSGWITPAGILLLASFFVERGILHRLMGNIGGF KDNRGEYGLTHDIFTTYCLKGSYSIRAQDHDAVMFRDILGYLSRVPTESFQRIKQPQIRKE GQLSERKTDKFITFALNYLEDYGLKDLEGCKACFARSKIVREQENVESINDKEYKPHENKK KVEIHFDQSKEDRFYINRNNVILKIQKKDGHSNIVRMGVYELKYLVLMSLVGKAKEAVEKI DNYIQDLRDQLPYIEGKNKEEIKEYVRFFPRFIRSHLGLLQINDEEKIKARLDYVKTKWLD KKEKSKELELHKKGRDILRYINERCDRELNRNVYNRILELLVSKDLTGFYRELEELKRTRR IDKNIVQNLSGQKTINALHEKVCDLVLKEIESLDTENLRKYLGLIPKEEKEVTFKEKVDRI LKQPVIYKGFLRYQFFKDDKKSFVLLVEDALKEKGGGCDVPLGKEYYKIVSLDKYDKENKT LCETLAMDRLCLMMARQYYLSLNAKLAQEAQQIEWKKEDSIELIIFTLKNPDQSKQSFSIR FSVRDFTKLYVTDDPEFLARLCSYFFPVEKEIEYHKLYSEGINKYTNLQKEGIEAILELEK KLIERNRIQSAKNYLSFNEIMNKSGYNKDEQDDLKKVRNSLLHYKLIFEKEHLKKFYEVMR GEGIEKKWSLIV* (SEQ ID NO: 2)
Cas13f.1	MNGIELKKEEAAFYFNQAELNLKAIEDNIFDKERRKTLLNNPQILAKMENFIFNFRDVTKN AKGEIDCLLLKLRELRNFYSHYVHKRDVRELSKGEKPILEKYYQFAIESTGSENVKLEIIE NDAWLADAGVLFFLCIFLKKSQANKLISGISGFKRNDDTGQPRRNLFTYFSIREGYKVVPE MQKHFLLFSLVNHLSNQDDYIEKAHQPYDIGEGLFFHRIASTFLNISGILRNMKFYTYQSK RLVEQRGELKREKDIFAWEEPFQGNSYFEINGHKGVIGEDELKELCYAFLIGNQDANKVEG RITQFLEKFRNANSVQQVKDDEMLKPEYFPANYFAESGVGRIKDRVLNRLNKAIKSNKAKK GEIIAYDKMREVMAFINNSLPVDEKLKPKDYKRYLGMVRFWDREKDNIKREFETKEWSKYL PSNFWTAKNLERVYGLAREKNAELFNKLKADVEKMDERELEKYQKINDAKDLANLRRLASD FGVKWEEKDWDEYSGQIKKQITDSQKLTIMKQRITAGLKKKHGIENLNLRITIDINKSRKA VLNRIAIPRGFVKRHILGWQESEKVSKKIREAECEILLSKEYEELSKQFFQSKDYDKMTRI NGLYEKNKLIALMAVYLMGQLRILFKEHTKLDDITKTTVDFKISDKVTVKIPFSNYPSLVY TMSSKYVDNIGNYGFSNKDKDKPILGKIDVIEKQRMEFIKEVLGFEKYLFDDKIIDKSKFA DTATHISFAEIVEELVEKGWDKDRLTKLKDARNKALHGEILTGTSFDETKSLINELKK* (SEQ ID NO: 3)
Cas13f.2	MSPDFIKLEKQEAAFYFNQTELNLKAIESNILDKQQRMILLNNPRILAKVGNFIFNFRDVT KNAKGEIDCLLFKLEELRNFYSHYVHTDNVKELSNGEKPLLERYYQIAIQATRSEDVKFEL FETRNENKITDAGVLFFLCMFLKKSQANKLISGISGFKRNDPTGQPRRNLFTYFSAREGYK ALPDMQKHFLLFTLVNYLSNQDEYISELKQYGEIGQGAFFNRIASTFLNISGISGNTKFYS YQSKRIKEQRGELNSEKDSFEWIEPFQGNSYFEINGHKGVIGEDELKELCYALLVAKQDIN AVEGKIMQFLKKFRNTGNLQQVKDDEMLEIEYFPASYFNESKKEDIKKEILGRLDKKIRSC SAKAEKAYDKMKEVMEFINNSLPAEEKLKRKDYRRYLKMVRFWSREKGNIEREFRTKEWSK YFSSDFWRKNNLEDVYKLATQKNAELFKNLKAAAEKMGETEFEKYQQINDVKDLASLRRLT QDFGLKWEEKDWEEYSEQIKKQITDRQKLTIMKQRVTAELKKKHGIENLNLRITIDSNKSR KAVLNRIAIPRGFVKKHILGWQGSEKISKNIREAECKILLSKKYEELSRQFFEAGNFDKLT QINGLYEKNKLTAFMSVYLMGRLNIQLNKHTELGNLKKTEVDFKISDKVTEKIPFSQYPSL VYAMSRKYVDNVDKYKFSHQDKKKPFLGKIDSIEKERIEFIKEVLDFEEYLFKNKVIDKSK FSDTATHISFKEICDEMGKKGCNRNKLTELNNARNAALHGEIPSETSFREAKPLINELKK* (SEQ ID NO: 4)

Cas13f.3	MSPDFIKLEKQEAAFYFNQTELNLKAIESNIFDKQQRVILLNNPQILAKVGDFIFNFRDVT KNAKGEIDCLLLKLRELRNFYSHYVYTDDVKILSNGERPLLEKYYQFAIEATGSENVKLEI IESNNRLTEAGVLFFLCMFLKKSQANKLISGISGFKRNDPTGQPRRNLFTYFSVREGYKVV PDMQKHFLLFVLVNHLSGQDDYIEKAQKPYDIGEGLFFHRIASTFLNISGILRNMEFYIYQ SKRLKEQQGELKREKDIFPWIEPFQGNSYFEINGNKGIIGEDELKELCYALLVAGKDVRAV EGKITQFLEKFKNADNAQQVEKDEMLDRNNFPANYFAESNIGSIKEKILNRLGKTDDSYNK TGTKIKPYDMMKEVMEFINNSLPADEKLKRKDYRRYLKMVRIWDSEKDNIKREFESKEWSK YFSSDFWMAKNLERVYGLAREKNAELFNKLKAVVEKMDEREFEKYRLINSAEDLASLRRLA KDFGLKWEEKDWQEYSGQIKKQISDRQKLTIMKQRITAELKKKHGIENLNLRITIDSNKSR KAVLNRIAVPRGFVKEHILGWQGSEKVSKKTREAKCKILLSKEYEELSKQFFQTRNYDKMT QVNGLYEKNKLLAFMVVYLMERLNILLNKPTELNELEKAEVDFKISDKVMAKIPFSQYPSL VYAMSSKYADSVGSYKFENDEKNKPFLGKIDTIEKQRMEFIKEVLGFEEYLFEKKIIDKSE FADTATHISFDEICNELIKKGWDKDKLTKLKDARNAALHGEIPAETSFREAKPLINGLKK* (SEQ ID NO: 5)
Cas13f.4	MNIIKLKKEEAAFYFNQTILNLSGLDEIIEKQIPHIISNKENAKKVIDKIFNNRLLLKSVE NYIYNFKDVAKNARTEIEAILLKLVELRNFYSHYVHNDTVKILSNGEKPILEKYYQIAIEA TGSKNVKLVIIENNNCLTDSGVLFLLCMFLKKSQANKLISSVSGFKRNDKEGQPRRNLFTY YSVREGYKVVPDMQKHFLLFALVNHLSEQDDHIEKQQQSDELGKGLFFHRIASTFLNESGI FNKMQFYTYQSNRLKEKRGELKHEKDTFTWIEPFQGNSYFTLNGHKGVISEDQLKELCYTI LIEKQNVDSLEGKIIQFLKKFQNVSSKQQVDEDELLKREYFPANYFGRAGTGTLKEKILNR LDKRMDPTSKVTDKAYDKMIEVMEFINMCLPSDEKLRQKDYRRYLKMVRFWNKEKHNIKRE FDSKKWTRFLPTELWNKRNLEEAYQLARKENKKKLEDMRNQVRSLKENDLEKYQQINYVND LENLRLLSQELGVKWQEKDWVEYSGQIKKQISDNQKLTIMKQRITAELKKMHGIENLNLRI SIDTNKSRQTVMNRIALPKGFVKNHIQQNSSEKISKRIREDYCKIELSGKYEELSRQFFDK KNFDKMTLINGLCEKNKLIAFMVIYLLERLGFELKEKTKLGELKQTRMTYKISDKVKEDIP LSYYPKLVYAMNRKYVDNIDSYAFAAYESKKAILDKVDIIEKQRMEFIKQVLCFEEYIFEN RIIEKSKFNDEETHISFTQIHDELIKKGRDTEKLSKLKHARNKALHGEIPDGTSFEKAKLL INEIKK* (SEQ ID NO: 6)
Cas13f.5	MNAIELKKEEAAFYFNQARLNISGLDEIIEKQLPHIGSNRENAKKTVDMILDNPEVLKKME NYVFNSRDIAKNARGELEALLLKLVELRNFYSHYVHKDDVKTLSYGEKPLLDKYYEIAIEA TGSKDVRLEIIDDKNKLTDAGVLFLLCMFLKKSEANKLISSIRGFKRNDKEGQPRRNLFTY YSVREGYKVVPDMQKHFLLFTLVNHLSNQDEYISNLRPNQEIGQGGFFHRIASKFLSDSGI LHSMKFYTYRSKRLTEQRGELKPKKDHFTWIEPFQGNSYFSVQGQKGVIGEEQLKELCYVL LVAREDFRAVEGKVTQFLKKFQNANNVQQVEKDEVLEKEYFPANYFENRDVGRVKDKILNR LKKITESYKAKGREVKAYDKMKEVMEFINNCLPTDENLKLKDYRRYLKMVRFWGREKENIK REFDSKKWERFLPRELWQKRNLEDAYQLAKEKNTELFNKLKTTVERMNELEFEKYQQINDA KDLANLRQLARDFGVKWEEKDWQEYSGQIKKQITDRQKLTIMKQRITAALKKKQGIENLNL RITTDTNKSRKVVLNRIALPKGFVRKHILKTDIKISKQIRQSQCPIILSNNYMKLAKEFFE ERNFDKMTQINGLFEKNVLIAFMIVYLMEQLNLRLGKNTELSNLKKTEVNFTITDKVTEKV QISQYPSLVFAINREYVDGISGYKLPPKKPKEPPYTFFEKIDAIEKERMEFIKQVLGFEEH LFEKNVIDKTRFTDTATHISFNEICDELIKKGWDENKIIKLKDARNAALHGKIPEDTSFDE AKVLINELKK* (SEQ ID NO: 7)

In the sequences above, the two RX4-6H (RXXXXH) motifs in each effector are double-underlined. In Cas13e.1, the C-terminal motif may have two possibilities due to the RR and HH sequences flanking the motif. Mutations at one or both such domains may create an RNase dead

version (or "dCas) of the Cas13e and Cas13f effector proteins, homologs, orthologs, fusions, conjugates, derivatives, or functional fragments thereof, while substantially maintaining their ability to bind the guide RNA and the target RNA complementary to the guide RNA.

The corresponding DR coding sequences for the Cas effectors are listed below:

Cas13e.1	GCTGGAGCAGCCCCCGATTTGTGGGGTGATTACAGC (SEQ ID NO: 8)
Cas13e.2	GCTGAAGAAGCCTCCGATTTGAGAGGTGATTACAGC (SEQ ID NO: 9)
Cas13f.1	GCTGTGATAGACCTCGATTTGTGGGGTAGTAACAGC (SEQ ID NO: 10)
Cas13f.2	GCTGTGATAGACCTCGATTTGTGGGGTAGTAACAGC (SEQ ID NO: 11)
Cas13f.3	GCTGTGATAGACCTCGATTTGTGGGGTAGTAACAGC (SEQ ID NO: 12)
Cas13f.4	GCTGTGATGGGCCTCAATTTGTGGGGAAGTAACAGC (SEQ ID NO: 13)
Cas13f.5	GCTGTGATAGGCCTCGATTTGTGGGGTAGTAACAGC (SEQ ID NO: 14)

Since the secondary structures of the DR sequences, including the location and size of the step, bulge, and loop structures, are likely more important than the specific nucleotide sequences that form such secondary structures, alternative or derivative DR sequences can also be used in the systems and methods of the invention, so long as these derivative or alternative DR sequences have a secondary structure that substantially resembles the secondary structure of an RNA encoded by any one of SEQ ID NO: 8-14. For example, the derivative DR sequence may have ± 1 or 2 base pair(s) in one or both stems (see FIG. 2), have ± 1, 2, or 3 bases in either or both of the single strands in the bulge, and/or have ± 1, 2, 3, or 4 bases in the loop region.

In some embodiments, a Type VI-E and VI-F CRISPR-Cas effector proteins include a "derivative" having an amino acid sequence with at least about 80% sequence identity to the amino acid sequence of any one of SEQ ID NOs: 1-7 above (*e.g.*, 81%, 82%, 83%, 84%, 85%, 86%, 87% 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100%). Such derivative Cas effectors sharing significant protein sequence identity to any one of SEQ ID NOs: 1-7 have retained at least one of the functions of the Cas of SEQ ID NOs: 1-7 (see below), such as the ability to bind to and form a complex with a crRNA comprising at least one of the DR sequences of SEQ ID NOs: 8-14. For example, a Cas13e.1 derivative may share 85% amino acid sequence identity to SEQ ID NO: 1, 2, 3, 4, 5, 6, or 7, respectively, and retains the ability to bind to and form a complex with a crRNA having a DR sequence of SEQ ID NO: 8, 9, 10, 11, 12, 13, or 14, respectively.

In some embodiments, the derivative comprises conserved amino acid residue substitutions. In some embodiments, the derivative comprises only conserved amino acid residue substitutions (*i.e.*, all amino acid substitutions in the derivative are conserved substitutions, and there is no substitution that is not conserved).

In some embodiments, the derivative comprises no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acid insertions or deletions into any one of the wild-type sequences of SEQ ID NOs: 1-7. The insertion and/or deletion maybe clustered together, or separated throughout the entire length of the sequences, so long as at least one of the functions of the wild-type sequence is preserved. Such functions may include the ability to bind the guide / crRNA, the RNase activity, the ability to bind to and/or cleave the target RNA complementary to the guide / crRNA. In some embodiments, the insertions and/or deletions are not present in the RXXXXH motifs, or within 5, 10, 15, or 20 residues from the RXXXXH motifs.

In some embodiments, the derivative has retained the ability to bind guide RNA / crRNA.

In some embodiments, the derivative has retained the guide / crRNA-activated RNase activity.

In some embodiments, the derivative has retained the ability to bind target RNA and/or cleave the target RNA in the presence of the bound guide / crRNA that is complementary in sequence to at least a portion of the target RNA.

In other embodiments, the derivative has completely or partially lost the guide / crRNA-activated RNase activity, due to, for example, mutations in one or more catalytic residues of the RNA-guided RNase. Such derivatives are sometimes referred to as dCas, such as dCas13e.1, *etc*.

Thus in certain embodiments, the derivative may be modified to have diminished nuclease / RNase activity, *e.g.*, nuclease inactivation of at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 97%, or 100% as compared with the counterpart wild type proteins. The nuclease activity can be diminished by several methods known in the art, *e.g.*, introducing mutations into the nuclease (catalytic) domains of the proteins. In some embodiments, catalytic residues for the nuclease activities are identified, and these amino acid residues can be substituted by different amino acid residues (*e.g.*, glycine or alanine) to diminish the nuclease activity. In some embodiments, the amino acid substitution is a conservative amino acid substitution. In some embodiments, the amino acid substitution is a non-conservative amino

acid substitution.

In some embodiments, the modification comprises one or more mutations (*e.g.*, amino acid deletions, insertions, or substitutions) in at least one HEPN domain. In some embodiments, there is one, two, three, four, five, six, seven, eight, nine, or more amino acid substitutions in at least one HEPN domain. For example, in some embodiments, the one or more mutations comprise a substitution (*e.g.*, an alanine substitution) at an amino acid residue corresponding to R84, H89, R739, H744, R740, H745 of SEQ ID NO: 1, or R97, H102, R770, H775 of SEQ ID NO: 2, or R77, H82, R764, H769 of SEQ ID NO: 3, or R79, H84, R766A, H771 of SEQ ID NO: 4, or R79, H84, R766, H771 of SEQ ID NO: 5, or R89, H94, R773, H778 of SEQ ID NO: 6, or R89, H94, R777, H782 of SEQ ID NO: 7.

In certain embodiments, the one or more mutations or the two or more mutations may be in a catalytically active domain of the effector protein comprising a HEPN domain, or a catalytically active domain which is homologous to a HEPN domain. In certain embodiments, the effector protein comprises one or more of the following mutations: R84A, H89A, R739A, H744A, R740A, H745A (wherein amino acid positions correspond to amino acid positions of Cas13e.1). The skilled person will understand that corresponding amino acid positions in different Cas13e and Cas13f proteins may be mutated to the same effect. In certain embodiments, one or more mutations abolish catalytic activity of the protein completely or partially (*e.g.* altered cleavage rate, altered specificity, *etc.*).

Other exemplary (catalytic) residue mutations include: R97A, H102A, R770A, H775A of Cas13e.2, or R77A, H82A, R764A, H769A of Cas13f.1, or R79A, H84A, R766A, H771A of Cas13f.2, or R79A, H84A, R766A, H771A of Cas13f.3, or R89A, H94A, R773A, H778A of Cas13f.4, or R89A, H94A, R777A, H782A of Cas13f.5. In certain embodiments, any of the R and/or H residues herein may be replaced not be A but by G, V, or I.

The presence of at least one of these mutations results in a derivative having reduced or diminished RNase activity as compared to the corresponding wild-type protein lacking the mutations.

In certain embodiments, the effector protein as described herein is a "dead" effector protein, such as a dead Cas13e or Cas13f effector protein (*i.e.* dCas13e and dCas13f). In certain embodiments, the effector protein has one or more mutations in HEPN domain 1 (N-terminal).

In certain embodiments, the effector protein has one or more mutations in HEPN domain 2 (C-terminal). In certain embodiments, the effector protein has one or more mutations in HEPN domain 1 and HEPN domain 2.

The inactivated Cas or derivative or functional fragment thereof can be fused or associated with one or more heterologous / functional domains (*e.g.*, via fusion protein, linker peptides, "GS" linkers, *etc.*). These functional domains can have various activities, *e.g.*, methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity, DNA cleavage activity, nucleic acid binding activity, base-editing activity, and switch activity (*e.g.*, light inducible). In some embodiments, the functional domains are Krüppel associated box (KRAB), SID (*e.g.* SID4X), VP64, VPR, VP16, Fok1, P65, HSF1, MyoD1, Adenosine Deaminase Acting on RNA such as ADAR1, ADAR2, APOBEC, cytidine deaminase (AID), TAD, mini-SOG, APEX, and biotin-APEX.

In some embodiments, the functional domain is a base editing domain, *e.g.*, ADAR1 (including wild-type or ADAR1DD version thereof, with or without the E1008Q), ADAR2 (including wild-type or ADAR2DD version thereof, with or without the E488Q mutation(s)), APOBEC, or AID.

In some embodiments, the functional domain may comprise one or more nuclear localization signal (NLS) domains. The one or more heterologous functional domains may comprise at least two or more NLS domains. The one or more NLS domain(s) may be positioned at or near or in proximity to a terminus of the effector protein (*e.g.*, Cas13e/Cas13f effector proteins) and if two or more NLSs, each of the two may be positioned at or near or in proximity to a terminus of the effector protein (*e.g.*, Cas13e/Cas13f effector proteins).

In some embodiments, at least one or more heterologous functional domains may be at or near the amino-terminus of the effector protein and/or wherein at least one or more heterologous functional domains is at or near the carboxy-terminus of the effector protein. The one or more heterologous functional domains may be fused to the effector protein. The one or more heterologous functional domains may be tethered to the effector protein. The one or more heterologous functional domains may be linked to the effector protein by a linker moiety.

In some embodiments, multiple (*e.g.*, two, three, four, five, six, seven, eight, or more)

identical or different functional domains are present.

In some embodiments, the functional domain (*e.g.*, a base editing domain) is further fused to an RNA-binding domain (*e.g.*, MS2).

In some embodiments, the functional domain is associated to or fused via a linker sequence (*e.g.*, a flexible linker sequence or a rigid linker sequence). Exemplary linker sequences and functional domain sequences are provided in table below.

Amino Acid Sequences of Motifs and Functional Domains in Engineered Variants of Type VI-E and VI-F CRISPR Cas Effectors

Linker 1	GS
Linker 2	GSGGGGS
Linker 3	GGGGSGGGGSGGGGS
ADAR1DD-WT	SLGTGNRCVKGDSLSLKGETVNDCHAEIISRRGFIRFLYSELMKYNSQTAKDSIFEPAKGG EKLQIKKTVSFHLYISTAPCGDGALFDKSCSDRAMESTESRHYPVFENPKQGKLRTKVENG EGTIPVESSDIVPTWDGIRLGERLRTMSCSDKILRWNVLGLQGALLTHFLQPIYLKSVTLG YLFSQGHLTRAICCRVTRDGSAFEDGLRHPFIVNHPKVGRVSIYDSKRQSGKTKETSVNWC LADGYDLEILDGTRGTVDGPRNELSRVSKKNIFLLFKKLCSFRYRRDLLRLSYGEAKKAAR DYETAKNYFKKGLKDMGYGNWISKPQEEKNF
ADAR1DD-E1008Q	SLGTGNRCVKGDSLSLKGETVNDCHAEIISRRGFIRFLYSELMKYNSQTAKDSIFEPAKGG EKLQIKKTVSFHLYISTAPCGDGALFDKSCSDRAMESTESRHYPVFENPKQGKLRTKVENG QGTIPVESSDIVPTWDGIRLGERLRTMSCSDKILRWNVLGLQGALLTHFLQPIYLKSVTLG YLFSQGHLTRAICCRVTRDGSAFEDGLRHPFIVNHPKVGRVSIYDSKRQSGKTKETSVNWC LADGYDLEILDGTRGTVDGPRNELSRVSKKNIFLLFKKLCSFRYRRDLLRLSYGEAKKAAR DYETAKNYFKKGLKDMGYGNWISKPQEEKNF
ADAR2DD-WT	QLHLPQVLADAVSRLVLGKFGDLTDNFSSPHARRKVLAGVVMTTGTDVKDAKVISVSTGTK CINGEYMSDRGLALNDCHAEIISRRSLLRFLYTQLELYLNNKDDQKRSIFQKSERGGFRLK ENVQFHLYISTSPCGDARIFSPHEPILEEPADRHPNRKARGQLRTKIESGEGTIPVRSNAS IQTWDGVLQGERLLTMSCSDKIARWNVVGIQGSLLSIFVEPIYFSSIILGSLYHGDHLSRA MYQRISNIEDLPPLYTLNKPLLSGISNAEARQPGKAPNFSVNWTVGDSAIEVINATTGKDE LGRASRLCKHALYCRWMRVHGKVPSHLLRSKITKPNVYHESKLAAKEYQAAKARLFTAFIK AGLGAWVEKPTEQDQFSLT
ADAR2DD-E488Q	QLHLPQVLADAVSRLVLGKFGDLTDNFSSPHARRKVLAGVVMTTGTDVKDAKVISVSTGTK CINGEYMSDRGLALNDCHAEIISRRSLLRFLYTQLELYLNNKDDQKRSIFQKSERGGFRLK ENVQFHLYISTSPCGDARIFSPHEPILEEPADRHPNRKARGQLRTKIESGQGTIPVRSNAS IQTWDGVLQGERLLTMSCSDKIARWNVVGIQGSLLSIFVEPIYFSSIILGSLYHGDHLSRA MYQRISNIEDLPPLYTLNKPLLSGISNAEARQPGKAPNFSVNWTVGDSAIEVINATTGKDE LGRASRLCKHALYCRWMRVHGKVPSHLLRSKITKPNVYHESKLAAKEYQAAKARLFTAFIK AGLGAWVEKPTEQDQFSLT
AID-APOBEC1	MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLF LRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAE PEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPL

	YEVDDLRDAFRTLGL
Lamprey_AID-APOBEC1	MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGT ERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIW ACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKR AEKRRSELSIMIQVKILHTTKSPAV
APOBEC1_BE1	MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRK*ETC*LLYEINWGGRHSIWRHTSQNTNKH VEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELY CIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK

The positioning of the one or more functional domains on the inactivated Cas proteins is one that allows for correct spatial orientation for the functional domain to affect the target with the attributed functional effect. For example, if the functional domain is a transcription activator (*e.g.*, VP16, VP64, or p65), the transcription activator is placed in a spatial orientation that allows it to affect the transcription of the target. Likewise, a transcription repressor is positioned to affect the transcription of the target, and a nuclease (*e.g.*, Fok1) is positioned to cleave or partially cleave the target. In some embodiments, the functional domain is positioned at the N-terminus of the Cas / dCas. In some embodiments, the functional domain is positioned at the C-terminus of the Cas / dCas. In some embodiments, the inactivated CRISPR-associated protein (dCas) is modified to comprise a first functional domain at the N-terminus and a second functional domain at the C-terminus.

Various examples of inactivated CRISPR-associated proteins fused with one or more functional domains and methods of using the same are described, *e.g.*, in International Publication No. WO 2017/219027, which is incorporated herein by reference in its entirety, and in particular with respect to the features described herein.

In some embodiments, a Type VI-E and VI-F CRISPR-Cas effector proteins includes the amino acid sequence of any one of SEQ ID NOs: 1-7 above. In some embodiments, a Type VI-E and VI-F CRISPR-Cas effector proteins excludes the naturally occurring amino acid sequence of any one of SEQ ID NOs: 1-7 above.

In some embodiments, instead of using full-length wild-type (SEQ ID NOs: 1-7) or derivative Type VI-E and VI-F Cas effectors, "functional fragments" thereof can be used.

A "functional fragment," as used herein, refers to a fragment of a wild-type protein of

any one of SEQ ID NOs: 1-7, or a derivative thereof, that has less-than full-length sequence. The deleted residues in the functional fragment can be at the N-terminus, the C-terminus, and/or internally. The functional fragment retains at least one function of the wild-type VI-E or VI-F Cas, or at least one function of its derivative. Thus a functional fragment is defined specifically with respect to the function at issue. For example, a functional fragment, wherein the function is the ability to bind crRNA and target RNA, may not be a functional fragment with respect to the RNase function, because losing the RXXXXH motifs at both ends of the Cas may not affect its ability to bind a crRNA and target RNA, but may eliminate destroy the RNase activity.

In some embodiments, compared to full-length sequences SEQ ID NOs: 1-7, the Type VI-E or VI-F CRISPR-Cas effector proteins or derivatives thereof or functional fragments thereof lacks about 30, 60, 90, 120, 150, or about 180 residues from the N-terminus.

In some embodiments, compared to full-length sequences SEQ ID NOs: 1-7, the Type VI-E or VI-F CRISPR-Cas effector proteins or derivatives thereof or functional fragments thereof lacks about 30, 60, 90, 120, or about 150 residues from the C-terminus.

In some embodiments, compared to full-length sequences SEQ ID NOs: 1-7, the Type VI-E or VI-F CRISPR-Cas effector proteins or derivatives thereof or functional fragments thereof lacks about 30, 60, 90, 120, 150, or about 180 residues from the N-terminus, and lacks about 30, 60, 90, 120, or about 150 residues from the C-terminus.

In some embodiments, the Type VI-E or VI-F CRISPR-Cas effector proteins or derivatives thereof or functional fragments thereof have RNase activity, *e.g.*, guide / crRNA-activated specific RNase activity.

In some embodiments, the Type VI-E or VI-F CRISPR-Cas effector proteins or derivatives thereof or functional fragments thereof have no substantial / detectable collateral RNase activity.

Here, "collateral RNase activity" refers to the non-specific RNase activity observed in certain other Class 2, type VI RNA-guided RNases, such as Cas13a. A complex comprising Cas13a, for example, upon activation by binding to a target nucleic acid (*e.g.*, a target RNA), a conformational change results, which in turn causes the complex to act as a non-specific RNase, cleaving and/or degrading nearby RNA molecules (*e.g.*, ssRNA or dsRNA molecules) (*i.e.*, "collateral" effects).

In certain embodiments, a complex comprised of (but not limited to) the Type VI-E or VI-F CRISPR-Cas effector proteins or derivatives thereof or functional fragments thereof and a crRNA does not exhibit collateral RNase activity subsequent to target recognition. This "collateral-free" embodiment may comprise wild-type, engineered / derivative effector proteins, or functional fragments thereof.

In some embodiments, the Type VI-E or VI-F CRISPR-Cas effector proteins or derivatives thereof or functional fragments thereof recognizes and cleaves the target RNA without any additional requirements adjacent to or flanking the protospacer (*i.e.*, protospacer adjacent motif "PAM" or protospacer flanking sequence "PFS" requirements).

The present disclosure also provides a split version of the CRISPR-associated proteins described herein (*e.g.*, a Type VI-E or VI-F CRISPR-Cas effector protein). The split version of the CRISPR-associated protein may be advantageous for delivery. In some embodiments, the CRISPR-associated proteins are split into two parts of the enzyme, which together substantially comprise a functioning CRISPR-associated protein.

The split can be done in a way that the catalytic domain(s) are unaffected. The CRISPR-associated protein may function as a nuclease or may be an inactivated enzyme, which is essentially a RNA-binding protein with very little or no catalytic activity (*e.g.*, due to mutation(s) in its catalytic domains). Split enzymes are described, *e.g.*, in Wright *et al.*, "Rational design of a split-Cas9 enzyme complex," *Proc. Nat'l. Acad. Sci.* 112(10): 2984-2989, 2015, which is incorporated herein by reference in its entirety.

For example, in some embodiments, the nuclease lobe and α-helical lobe are expressed as separate polypeptides. Although the lobes do not interact on their own, the crRNA recruits them into a ternary complex that recapitulates the activity of full-length CRISPR-associated proteins and catalyzes site-specific DNA cleavage. The use of a modified crRNA abrogates split-enzyme activity by preventing dimerization, allowing for the development of an inducible dimerization system.

In some embodiments, the split CRISPR-associated protein can be fused to a dimerization partner, *e.g.*, by employing rapamycin sensitive dimerization domains. This allows the generation of a chemically inducible CRISPR-associated protein for temporal control of the activity of the protein. The CRISPR-associated protein can thus be rendered chemically

inducible by being split into two fragments and rapamycin-sensitive dimerization domains can be used for controlled re-assembly of the protein.

The split point is typically designed *in silico* and cloned into the constructs. During this process, mutations can be introduced to the split CRISPR-associated protein and non-functional domains can be removed.

In some embodiments, the two parts or fragments of the split CRISPR-associated protein (*i.e.*, the N-terminal and C-terminal fragments), can form a full CRISPR-associated protein, comprising, *e.g.*, at least 70%, at least 80%, at least 90%, at least 95%, or at least 99% of the sequence of the wild-type CRISPR-associated protein.

The CRISPR-associated proteins described herein (*e.g.*, a Type VI-E or VI-F CRISPR-Cas effector protein) can be designed to be self-activating or self-inactivating. For example, the target sequence can be introduced into the coding construct of the CRISPR-associated protein. Thus, the CRISPR-associated protein can cleave the target sequence, as well as the construct encoding the protein thereby self-inactivating their expression. Methods of constructing a self-inactivating CRISPR system are described, *e.g.*, in Epstein and Schaffer, *Mol. Ther.* 24: S50, 2016, which is incorporated herein by reference in its entirety.

In some other embodiments, an additional crRNA, expressed under the control of a weak promoter (*e.g.*, 7SK promoter), can target the nucleic acid sequence encoding the CRISPR-associated protein to prevent and/or block its expression (*e.g.*, by preventing the transcription and/or translation of the nucleic acid). The transfection of cells with vectors expressing the CRISPR-associated protein, the crRNAs, and crRNAs that target the nucleic acid encoding the CRISPR-associated protein can lead to efficient disruption of the nucleic acid encoding the CRISPR-associated protein and decrease the levels of CRISPR-associated protein, thereby limiting the genome editing activity.

In some embodiments, the genome editing activity of the CRISPR-associated protein can be modulated through endogenous RNA signatures (*e.g.*, miRNA) in mammalian cells. A CRISPR-associated protein switch can be made by using a miRNA-complementary sequence in the 5'-UTR of mRNA encoding the CRISPR-associated protein. The switches selectively and efficiently respond to miRNA in the target cells. Thus, the switches can differentially control the genome editing by sensing endogenous miRNA activities within a heterogeneous cell population.

Therefore, the switch systems can provide a framework for cell-type selective genome editing and cell engineering based on intracellular miRNA information (see, *e.g.*, Hirosawa *et al.*, *Nucl. Acids Res.* 45(13): e118, 2017).

The CRISPR-associated proteins (*e.g.*, Type VI-E and VI-F CRISPR-Cas effector proteins) can be inducibly expressed, *e.g.*, their expression can be light-induced or chemically-induced. This mechanism allows for activation of the functional domain in the CRISPR-associated proteins. Light inducibility can be achieved by various methods known in the art, *e.g.*, by designing a fusion complex wherein CRY2 PHR/CIBN pairing is used in split CRISPR-associated proteins (see, *e.g.*, Konermann *et al.*, "Optical control of mammalian endogenous transcription and epigenetic states," *Nature* 500:7463, 2013.

Chemical inducibility can be achieved, *e.g.*, by designing a fusion complex wherein FKBP/FRB (FK506 binding protein/FKBP rapamycin binding domain) pairing is used in split CRISPR-associated proteins. Rapamycin is required for forming the fusion complex, thereby activating the CRISPR-associated proteins (see, *e.g.*, Zetsche *et al.*, "A split-Cas9 architecture for inducible genome editing and transcription modulation," *Nature Biotech.* 33:2:139-42, 2015).

Furthermore, expression of the CRISPR-associated proteins can be modulated by inducible promoters, *e.g.*, tetracycline or doxycycline controlled transcriptional activation (Tet-On and Tet-Off expression system), hormone inducible gene expression system (*e.g.*, an ecdysone inducible gene expression system), and an arabinose-inducible gene expression system. When delivered as RNA, expression of the RNA targeting effector protein can be modulated via a riboswitch, which can sense a small molecule like tetracycline (see, *e.g.*, Goldfless *et al.*, "Direct and specific chemical control of eukaryotic translation with a synthetic RNA-protein interaction," *Nucl. Acids Res.* 40:9: e64-e64, 2012).

Various embodiments of inducible CRISPR-associated proteins and inducible CRISPR systems are described, *e.g.*, in U.S. Pat. No. 8,871,445, US Publication No. 2016/0208243, and International Publication No. WO 2016/205764, each of which is incorporated herein by reference in its entirety.

In some embodiments, the CRISPR-associated proteins include at least one (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) Nuclear Localization Signal (NLS) attached to the N-terminal or C-terminal of the protein. Non-limiting examples of NLSs include an NLS sequence derived from:

the NLS of the SV40 virus large T-antigen, having the amino acid sequence PKKKRKV; the NLS from nucleoplasmin (*e.g.*, the nucleoplasmin bipartite NLS with the sequence KRPAATKKAGQAKKKK); the c-myc NLS having the amino acid sequence PAAKRVKLD or RQRRNELKRSP; the hRNPA1 M9 NLS having the sequence NQSSNFGPMKGGNFGGRSS GPYGGGGQYFAKPRNQGGY; the sequence RMRIZFKNKGKDTAELRRRRVEVSVELRK AKKDEQILKRRNV of the IBB domain from importin-alpha; the sequences VSRKRPRP and PPKKARED of the myoma T protein; the sequence PQPKKKPL of human p53; the sequence SALIKKKKKMAP of mouse c-abl IV; the sequences DRLRR and PKQKKRK of the influenza virus NS1; the sequence RKLKKKIKKL of the Hepatitis virus delta antigen; the sequence REKKKFLKRR of the mouse Mx1 protein; the sequence KRKGDEVDGVDEVAKKKSKK of the human poly(ADP-ribose) polymerase; and the sequence RKCLQAGMNLEARKTKK of the human glucocorticoid receptor. In some embodiments, the CRISPR-associated protein comprises at least one (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) Nuclear Export Signal (NES) attached the N-terminal or C-terminal of the protein. In a preferred embodiment a C-terminal and/or N-terminal NLS or NES is attached for optimal expression and nuclear targeting in eukaryotic cells, *e.g.*, human cells.

In some embodiments, the CRISPR-associated proteins described herein are mutated at one or more amino acid residues to alter one or more functional activities.

For example, in some embodiments, the CRISPR-associated protein is mutated at one or more amino acid residues to alter its helicase activity.

In some embodiments, the CRISPR-associated protein is mutated at one or more amino acid residues to alter its nuclease activity (*e.g.*, endonuclease activity or exonuclease activity).

In some embodiments, the CRISPR-associated protein is mutated at one or more amino acid residues to alter its ability to functionally associate with a guide RNA.

In some embodiments, the CRISPR-associated protein is mutated at one or more amino acid residues to alter its ability to functionally associate with a target nucleic acid.

In some embodiments, the CRISPR-associated proteins described herein are capable of cleaving a target RNA molecule.

In some embodiments, the CRISPR-associated protein is mutated at one or more amino acid residues to alter its cleaving activity. For example, in some embodiments, the CRISPR-

associated protein may comprise one or more mutations that render the enzyme incapable of cleaving a target nucleic acid.

In some embodiments, the CRISPR-associated protein is capable of cleaving the strand of the target nucleic acid that is complementary to the strand to which the guide RNA hybridizes.

In some embodiments, a CRISPR-associated protein described herein can be engineered to have a deletion in one or more amino acid residues to reduce the size of the enzyme while retaining one or more desired functional activities (*e.g.*, nuclease activity and the ability to interact functionally with a guide RNA). The truncated CRISPR-associated protein can be advantageously used in combination with delivery systems having load limitations.

In some embodiments, the CRISPR-associated proteins described herein can be fused to one or more peptide tags, including a His-tag, GST-tag, a V5-tag, FLAG-tag, HA-tag, VSV-G-tag, Trx-tag, or myc-tag.

In some embodiments, the CRISPR-associated proteins described herein can be fused to a detectable moiety such as GST, a fluorescent protein (*e.g.*, GFP, HcRed, DsRed, CFP, YFP, or BFP), or an enzyme (such as HRP or CAT).

In some embodiments, the CRISPR-associated proteins described herein can be fused to MBP, LexA DNA binding domain, or Gal4 DNA-binding domain.

In some embodiments, the CRISPR-associated proteins described herein can be linked to or conjugated with a detectable label such as a fluorescent dye, including FITC and DAPI.

In any of the embodiments herein, the linkage between the CRISPR-associated proteins described herein and the other moiety can be at the N- or C-terminal of the CRISPR-associated proteins, and sometimes even internally via covalent chemical bonds. The linkage can be effected by any chemical linkage known in the art, such as peptide linkage, linkage through the side chain of amino acids such as D, E, S, T, or amino acid derivatives (Ahx, β-Ala, GABA or Ava), or PEG linkage.

3. Polynucleotides

The invention also provides nucleic acids encoding the proteins and guide RNAs (*e.g.*, a crRNA) described herein (*e.g.*, a CRISPR-associated protein or an accessory protein).

In some embodiments, the nucleic acid is a synthetic nucleic acid. In some embodiments, the nucleic acid is a DNA molecule. In some embodiments, the nucleic acid is an RNA molecule (*e.g.*, an mRNA molecule encoding the Cas, derivative or functional fragment thereof). In some embodiments, the mRNA is capped, polyadenylated, substituted with 5-methyl cytidine, substituted with pseudouridine, or a combination thereof.

In some embodiments, the nucleic acid (*e.g.*, DNA) is operably linked to a regulatory element (*e.g.*, a promoter) in order to control the expression of the nucleic acid. In some embodiments, the promoter is a constitutive promoter. In some embodiments, the promoter is an inducible promoter. In some embodiments, the promoter is a cell-specific promoter. In some embodiments, the promoter is an organism-specific promoter.

Suitable promoters are known in the art and include, for example, a pol I promoter, a pol II promoter, a pol III promoter, a T7 promoter, a U6 promoter, a H1 promoter, retroviral Rous sarcoma virus LTR promoter, a cytomegalovirus (CMV) promoter, a SV40 promoter, a dihydrofolate reductase promoter, and a β-actin promoter. For example, a U6 promoter can be used to regulate the expression of a guide RNA molecule described herein.

In some embodiments, the nucleic acid(s) are present in a vector (*e.g.*, a viral vector or a phage). The vector can be a cloning vector, or an expression vector. The vectors can be plasmids, phagemids, Cosmids, *etc.* The vectors may include one or more regulatory elements that allow for the propagation of the vector in a cell of interest (*e.g.*, a bacterial cell or a mammalian cell). In some embodiments, the vector includes a nucleic acid encoding a single component of a CRISPR-associated (Cas) system described herein. In some embodiments, the vector includes multiple nucleic acids, each encoding a component of a CRISPR-associated (Cas) system described herein.

In one aspect, the present disclosure provides nucleic acid sequences that are at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the nucleic acid sequences described herein, *i.e.*, nucleic acid sequences encoding the Cas proteins, derivatives, functional fragments, or guide / crRNA, including the DR sequences of SEQ ID NOs: 8-14.

In another aspect, the present disclosure also provides nucleic acid sequences encoding amino acid sequences that are at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%,

92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequences described herein, such as SEQ ID NOs: 1-7.

In some embodiments, the nucleic acid sequences have at least a portion (*e.g.*, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotides, *e.g.*, contiguous or non-contiguous nucleotides) that is the same as the sequences described herein. In some embodiments, the nucleic acid sequences have at least a portion (*e.g.*, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotides, *e.g.*, contiguous or non-contiguous nucleotides) that is different from the sequences described herein.

In related embodiments, the invention provides amino acid sequences having at least a portion (*e.g.*, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 30, 40, 50, 60, 70, 80, 90, or 100 amino acid residues, *e.g.*, contiguous or non-contiguous amino acid residues) that is the same as the sequences described herein. In some embodiments, the amino acid sequences have at least a portion (*e.g.*, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 30, 40, 50, 60, 70, 80, 90, or 100 amino acid residues, *e.g.*, contiguous or non-contiguous amino acid residues) that is different from the sequences described herein.

To determine the percent identity of two amino acid sequences, or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In general, the length of a reference sequence aligned for comparison purposes should be at least 80% of the length of the reference sequence, and in some embodiments is at least 90%, 95%, or 100% of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences. For purposes of the present disclosure, the comparison of sequences and determination of percent identity between two sequences can be accomplished using a Blossum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

The proteins described herein (*e.g.*, CRISPR-associated proteins or accessory proteins) can be delivered or used as either nucleic acid molecules or polypeptides.

In certain embodiments, the nucleic acid molecule encoding the CRISPR-associated proteins, derivatives or functional fragments thereof are codon-optimized for expression in a host cell or organism. The host cell may include established cell lines (such as 293T cells) or isolated primary cells. The nucleic acid can be codon optimized for use in any organism of interest, in particular human cells or bacteria. For example, the nucleic acid can be codon-optimized for any prokaryotes (such as *E. coli*), or any eukaryotes such as human and other non-human eukaryotes including yeast, worm, insect, plants and algae (including food crop, rice, corn, vegetables, fruits, trees, grasses), vertebrate, fish, non-human mammal (*e.g.*, mice, rats, rabbits, dogs, birds (such as chicken), livestock (cow or cattle, pig, horse, sheep, goat *etc.*), or non-human primates). Codon usage tables are readily available, for example, at the "Codon Usage Database" available at www.kazusa.orjp/codon/, and these tables can be adapted in a number of ways. See Nakamura *et al.*, *Nucl. Acids Res.* 28:292, 2000 (incorporated herein by reference in its entirety). Computer algorithms for codon optimizing a particular sequence for expression in a particular host cell are also available, such as Gene Forge (Aptagen; Jacobus, Pa.).

An example of a codon optimized sequence, is in this instance a sequence optimized for expression in a eukaryote, *e.g.*, humans (*i.e.* being optimized for expression in humans), or for another eukaryote, animal or mammal as herein discussed; see, *e.g.*, SaCas9 human codon optimized sequence in WO 2014/093622 (PCT/US2013/074667). Whilst this is preferred, it will be appreciated that other examples are possible and codon optimization for a host species other than human, or for codon optimization for specific organs is known. In general, codon optimization refers to a process of modifying a nucleic acid sequence for enhanced expression in the host cells of interest by replacing at least one codon (*e.g.* about or more than about 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more codons) of the native sequence with codons that are more frequently or most frequently used in the genes of that host cell while maintaining the native amino acid sequence. Various species exhibit particular bias for certain codons of a particular amino acid. Codon bias (differences in codon usage between organisms) often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, among other things, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs

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

4. RNA Guides or crRNA

In some embodiments, the CRISPR systems described herein include at least RNA guide (*e.g.*, a gRNA or a crRNA).

The architecture of multiple RNA guides is known in the art (see, *e.g.*, International Publication Nos. WO 2014/093622 and WO 2015/070083, the entire contents of each of which are incorporated herein by reference).

In some embodiments, the CRISPR systems described herein include multiple RNA guides (*e.g.*, one, two, three, four, five, six, seven, eight, or more RNA guides).

In some embodiments, the RNA guide includes a crRNA. In some embodiments, the RNA guide includes a crRNA but not a tracrRNA.

Sequences for guide RNAs from multiple CRISPR systems are generally known in the art, see, for example, Grissa *et al.* (*Nucleic Acids Res.* 35 (web server issue): W52-7, 2007; Grissa *et al.*, BMC Bioinformatics 8:172, 2007; Grissa *et al.*, Nucleic Acids Res. 36 (web server issue): W145-8, 2008; and Moller and Liang, PeerJ 5: e3788, 2017; the CRISPR database at: crispr.i2bc.paris-saclayfr/crispr/BLAST/CRISPRsBlast.php; and MetaCRAST available at: github.com/molleraj/MetaCRAST). All incorporated herein by reference.

In some embodiments, the crRNA includes a direct repeat (DR) sequence and a spacer

sequence. In certain embodiments, the crRNA comprises, consists essentially of, or consists of a direct repeat sequence linked to a guide sequence or spacer sequence, preferably at the 3'-end of the spacer sequence.

In general, the Cas protein forms a complex with the mature crRNA, which spacer sequence directs the complex to a sequence-specific binding with the target RNA that is complementary to the spacer sequence, and/or hybridizes to the spacer sequence. The resulting complex comprises the Cas protein and the mature crRNA bound to the target RNA.

The direct repeat sequences for the Cas13e and Cas13f systems are generally well conserved, especially at the ends, with a GCTG for Cas13e and GCTGT for Cas13f at the 5'-end, reverse complementary to a CAGC for Cas13e and ACAGC for Cas13f at the 3' end. This conservation suggests strong base pairing for an RNA stem-loop structure that potentially interacts with the protein(s) in the locus.

In some embodiments, the direct repeat sequence, when in RNA, comprises the general secondary structure of 5'-S1a-Ba-S2a-L-S2b-Bb-S1b-3', wherein segments S1a and S1b are reverse complement sequences and form a first stem (S1) having 4 nucleotides in Cas13e and 5 nucleotides in Cas13f; segments Ba and Bb do not base pair with each other and form a symmetrical or nearly symmetrical bulge (B), and have 5 nucleotides each in Cas13e, and 5 (Ba) and 4 (Bb) or 6 (Ba) and 5 (Bb) nucleotides respectively in Cas13f; segments S2a and S2b are reverse complement sequences and form a second stem (S2) having 5 base pairs in Cas13e and either 6 or 5 base pairs in Cas13f; and L is an 8-nucleotide loop in Cas13e and a 5-nucleotide loop in Cas13f. See FIG. 2.

In certain embodiments, S1a has a sequence of GCUG in Cas13e and GCUGU in Cas13f.

In certain embodiments, S2a has a sequence of GCCCC in Cas13e and A/G CCUC G/A in Cas13f (wherein the first A or G may be absent).

In some embodiments, the direct repeat sequence comprises or consists of a nucleic acid sequence of SEQ ID NOs: 8-14.

As used herein, "direct repeat sequence" may refer to the DNA coding sequence in the CRISPR locus, or to the RNA encoded by the same in crRNA. Thus when any of SEQ ID NOs: 8-14 is referred to in the context of an RNA molecule, such as crRNA, each T is understood to represent a U.

In some embodiments, the direct repeat sequence comprises or consists of a nucleic acid sequence having up to 1, 2, 3, 4, 5, 6, 7, or 8 nucleotides of deletion, insertion, or substitution of SEQ ID NOs: 8-14. In some embodiments, the direct repeat sequence comprises or consists of a nucleic acid sequence having at least 80%, 85%, 90%, 95%, or 97% of sequence identity with SEQ ID NOs: 8-14 (*e.g.*, due to deletion, insertion, or substitution of nucleotides in SEQ ID NOs: 8-14). In some embodiments, the direct repeat sequence comprises or consists of a nucleic acid sequence that is not identical to any one of SEQ ID NOs: 8-14, but can hybridize with a complement of any one of SEQ ID NOs: 8-14 under stringent hybridization conditions, or can bind to a complement of any one of SEQ ID NOs: 8-14 under physiological conditions.

In certain embodiments, the deletion, insertion, or substitution does not change the overall secondary structure of that of SEQ ID NOs: 8-14 (*e.g.*, the relative locations and/or sizes of the stems and bulges and loop do not significantly deviate from that of the original stems, bulges, and loop). For example, the deletion, insert, or substitution may be in the bulge or loop region so that the overall symmetry of the bulge remains largely the same. The deletion, insertion, or substitution may be in the stems so that the length of the stems do not significantly deviate from that of the original stems (*e.g.*, adding or deleting one base pair in each of the two stems correspond to 4 total base changes).

In certain embodiments, the deletion, insertion, or substitution results in a derivative DR sequence that may have ± 1 or 2 base pair(s) in one or both stems (see FIG. 2), have ± 1, 2, or 3 bases in either or both of the single strands in the bulge, and/or have ± 1, 2, 3, or 4 bases in the loop region.

In certain embodiments, any of the above direct repeat sequences that is different from any one of SEQ ID NOs: 8-14 retains the ability to function as a direct repeat sequence in the Cas13e or Cas13f proteins, as the DR sequence of SEQ ID NOs: 8-14.

In some embodiments, the direct repeat sequence comprises or consists of a nucleic acid having a nucleic acid sequence of any one of SEQ ID NOs: 8-14, with a truncation of the initial three, four, five, six, seven, or eight 3' nucleotides.

In some embodiments, the Cas protein comprises the amino acid sequence of SEQ ID NO: 1 and the crRNA comprises a direct repeat sequence, wherein the direct repeat sequence comprises or consists of the nucleic acid sequence of SEQ ID NO: 8.

In some embodiments, the Cas protein comprises the amino acid sequence of SEQ ID NO: 2 and the crRNA comprises a direct repeat sequence, wherein the direct repeat sequence comprises or consists of the nucleic acid sequence of SEQ ID NO: 9.

In some embodiments, the Cas protein comprises the amino acid sequence of SEQ ID NO: 3 and the crRNA comprises a direct repeat sequence, wherein the direct repeat sequence comprises or consists of the nucleic acid sequence of SEQ ID NO: 10.

In some embodiments, the Cas protein comprises the amino acid sequence of SEQ ID NO: 4 and the crRNA comprises a direct repeat sequence, wherein the direct repeat sequence comprises or consists of the nucleic acid sequence of SEQ ID NO: 11.

In some embodiments, the Cas protein comprises the amino acid sequence of SEQ ID NO: 5 and the crRNA comprises a direct repeat sequence, wherein the direct repeat sequence comprises or consists of the nucleic acid sequence of SEQ ID NO: 12.

In some embodiments, the Cas protein comprises the amino acid sequence of SEQ ID NO: 6 and the crRNA comprises a direct repeat sequence, wherein the direct repeat sequence comprises or consists of the nucleic acid sequence of SEQ ID NO: 13.

In some embodiments, the Cas protein comprises the amino acid sequence of SEQ ID NO: 7 and the crRNA comprises a direct repeat sequence, wherein the direct repeat sequence comprises or consists of the nucleic acid sequence of SEQ ID NO: 14.

In classic CRISPR systems, the degree of complementarity between a guide sequence (*e.g.*, a crRNA) and its corresponding target sequence can be about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or 100%. In some embodiments, the degree of complementarity is 90-100%.

The guide RNAs can be about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, 100, 125, 150, 175, 200 or more nucleotides in length. For example, for use in a functional Cas13e or Cas13f effector protein, or homologs, orthologs, derivatives, fusions, conjugates, or functional fragment thereof, the spacer can be between 10-60 nucleotides, 20-50 nucleotides, 25-45 nucleotides, 25-35 nucleotides, or about 27, 28, 29, 30, 31, 32, or 33 nucleotides. For use in dCas version of any of the above, however, the spacer can be between 10-200 nucleotides, 20-150 nucleotides, 25-100 nucleotides, 25-85 nucleotides, 35-75 nucleotides, 45-60 nucleotides, or about 46, 47, 48, 49, 50, 51, 52, 53, 54, or 55 nucleotides.

To reduce off-target interactions, *e.g.*, to reduce the guide interacting with a target sequence having low complementarity, mutations can be introduced to the CRISPR systems so that the CRISPR systems can distinguish between target and off-target sequences that have greater than 80%, 85%, 90%, or 95% complementarity. In some embodiments, the degree of complementarity is from 80% to 95%, *e.g.*, about 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, or 95% (for example, distinguishing between a target having 18 nucleotides from an off-target of 18 nucleotides having 1, 2, or 3 mismatches). Accordingly, in some embodiments, the degree of complementarity between a guide sequence and its corresponding target sequence is greater than 94.5%, 95%, 95.5%, 96%, 96.5%, 97%, 97.5%, 98%, 98.5%, 99%, 99.5%, or 99.9%. In some embodiments, the degree of complementarity is 100%.

It is known in the field that complete complementarity is not required, provided there is sufficient complementarity to be functional. Modulations of cleavage efficiency can be exploited by introduction of mismatches, *e.g.*, one or more mismatches, such as 1 or 2 mismatches between spacer sequence and target sequence, including the position of the mismatch along the spacer/target. The more central (*i.e.*, not at the 3' or 5'-ends) a mismatch, *e.g.*, a double mismatch, is located; the more cleavage efficiency is affected. Accordingly, by choosing mismatch positions along the spacer sequence, cleavage efficiency can be modulated. For example, if less than 100% cleavage of targets is desired (*e.g.*, in a cell population), 1 or 2 mismatches between spacer and target sequence can be introduced in the spacer sequences.

Type VI CRISPR-Cas effectors have been demonstrated to employ more than one RNA guide, thus enabling the ability of these effectors, and systems and complexes that include them, to target multiple nucleic acids. In some embodiments, the CRISPR systems described herein include multiple RNA guides (*e.g.*, two, three, four, five, six, seven, eight, nine, ten, fifteen, twenty, thirty, forty, or more) RNA guides. In some embodiments, the CRISPR systems described herein include a single RNA strand or a nucleic acid encoding a single RNA strand, wherein the RNA guides are arranged in tandem. The single RNA strand can include multiple copies of the same RNA guide, multiple copies of distinct RNA guides, or combinations thereof. The processing capability of the Type VI-E and VI-F CRISPR-Cas effector proteins described herein enables these effectors to be able to target multiple target nucleic acids (*e.g.*, target RNAs) without a loss of activity. In some embodiments, the Type VI-E and VI-F CRISPR-Cas effector

proteins may be delivered in complex with multiple RNA guides directed to different target RNA. In some embodiments, the Type VI-E and VI-F CRISPR-Cas effector proteins may be co-delivered with multiple RNA guides, each specific for a different target nucleic acid. Methods of multiplexing using CRISPR-associated proteins are described, for example, in U.S. Pat. No. 9,790,490 B2, and EP 3009511 B1, the entire contents of each of which are expressly incorporated herein by reference.

The spacer length of crRNAs can range from about 10-60 nucleotides, such as 15-50 nucleotides, 20-50 nucleotides, 25-50 nucleotide, or 19-50 nucleotides. In some embodiments, the spacer length of a guide RNA is at least 16 nucleotides, at least 17 nucleotides, at least 18 nucleotides, at least 19 nucleotides, at least 20 nucleotides, at least 21 nucleotides, or at least 22 nucleotides. In some embodiments, the spacer length is from 15 to 17 nucleotides (*e.g.*, 15, 16, or 17 nucleotides), from 17 to 20 nucleotides (*e.g.*, 17, 18, 19, or 20 nucleotides), from 20 to 24 nucleotides (*e.g.*, 20, 21, 22, 23, or 24 nucleotides), from 23 to 25 nucleotides (*e.g.*, 23, 24, or 25 nucleotides), from 24 to 27 nucleotides, from 27 to 30 nucleotides, from 30 to 45 nucleotides (*e.g.*, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, or 45 nucleotides), from 30 or 35 to 40 nucleotides, from 41 to 45 nucleotides, from 45 to 50 nucleotides (*e.g.*, 45, 46, 47, 48, 49, or 50 nucleotides), or longer. In some embodiments, the spacer length is from about 15 to about 42 nucleotides.

In some embodiments, the direct repeat length of the guide RNA is 15-36 nucleotides, is at least 16 nucleotides, is from 16 to 20 nucleotides (*e.g.*, 16, 17, 18, 19, or 20 nucleotides), is from 20-30 nucleotides (*e.g.*, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides), is from 30-40 nucleotides (*e.g.*, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 nucleotides), or is about 36 nucleotides (*e.g.*, 33, 34, 35, 36, 37, 38, or 39 nucleotides). In some embodiments, the direct repeat length of the guide RNA is 36 nucleotides.

In some embodiments, the overall length of the crRNA / guide RNA is about 36 nucleotides longer than any one of the spacer sequence length described herein above. For example, the overall length of the crRNA / guide RNA may be between 45-86 nucleotides, or 60-86 nucleotides, 62-86 nucleotides, or 63-86 nucleotides.

The crRNA sequences can be modified in a manner that allows for formation of a complex between the crRNA and CRISPR-associated protein and successful binding to the target, while at the same time not allowing for successful nuclease activity (*i.e.*, without nuclease

activity/without causing indels). These modified guide sequences are referred to as "dead crRNAs," "dead guides," or "dead guide sequences." These dead guides or dead guide sequences may be catalytically inactive or conformationally inactive with regard to nuclease activity. Dead guide sequences are typically shorter than respective guide sequences that result in active RNA cleavage. In some embodiments, dead guides are 5%, 10%, 20%, 30%, 40%, or 50%, shorter than respective guide RNAs that have nuclease activity. Dead guide sequences of guide RNAs can be from 13 to 15 nucleotides in length (*e.g.*, 13, 14, or 15 nucleotides in length), from 15 to 19 nucleotides in length, or from 17 to 18 nucleotides in length (*e.g.*, 17 nucleotides in length).

Thus, in one aspect, the disclosure provides non-naturally occurring or engineered CRISPR systems including a functional CRISPR-associated protein as described herein, and a crRNA, wherein the crRNA comprises a dead crRNA sequence whereby the crRNA is capable of hybridizing to a target sequence such that the CRISPR system is directed to a genomic locus of interest in a cell without detectable nuclease activity (*e.g.*, RNase activity).

A detailed description of dead guides is described, *e.g.*, in International Publication No. WO 2016/094872, which is incorporated herein by reference in its entirety.

Guide RNAs (*e.g.*, crRNAs) can be generated as components of inducible systems. The inducible nature of the systems allows for spatio-temporal control of gene editing or gene expression. In some embodiments, the stimuli for the inducible systems include, *e.g.*, electromagnetic radiation, sound energy, chemical energy, and/or thermal energy.

In some embodiments, the transcription of guide RNA (*e.g.*, crRNA) can be modulated by inducible promoters, *e.g.*, tetracycline or doxycycline controlled transcriptional activation (Tet-On and Tet-Off expression systems), hormone inducible gene expression systems (*e.g.*, ecdysone inducible gene expression systems), and arabinose-inducible gene expression systems. Other examples of inducible systems include, *e.g.*, small molecule two-hybrid transcription activations systems (FKBP, ABA, *etc.*), light inducible systems (Phytochrome, LOV domains, or cryptochrome), or Light Inducible Transcriptional Effector (LITE). These inducible systems are described, *e.g.*, in WO 2016205764 and U.S. Pat. No. 8,795,965, both of which are incorporated herein by reference in the entirety.

Chemical modifications can be applied to the crRNA's phosphate backbone, sugar, and/or base. Backbone modifications such as phosphorothioates modify the charge on the

phosphate backbone and aid in the delivery and nuclease resistance of the oligonucleotide (see, *e.g.*, Eckstein, "Phosphorothioates, essential components of therapeutic oligonucleotides," *Nucl. Acid Ther.*, 24, pp. 374-387, 2014); modifications of sugars, such as 2'-O-methyl (2'-OMe), 2'-F, and locked nucleic acid (LNA), enhance both base pairing and nuclease resistance (see, *e.g.*, Allerson *et al.* "Fully 2'-modified oligonucleotide duplexes with improved *in vitro* potency and stability compared to unmodified small interfering RNA," *J. Med. Chem.* 48.4: 901-904, 2005). Chemically modified bases such as 2-thiouridine or N6-methyladenosine, among others, can allow for either stronger or weaker base pairing (see, *e.g.*, Bramsen *et al.*, "Development of therapeutic-grade small interfering RNAs by chemical engineering," Front. Genet., 2012 Aug. 20; 3:154). Additionally, RNA is amenable to both 5' and 3' end conjugations with a variety of functional moieties including fluorescent dyes, polyethylene glycol, or proteins.

A wide variety of modifications can be applied to chemically synthesized crRNA molecules. For example, modifying an oligonucleotide with a 2'-OMe to improve nuclease resistance can change the binding energy of Watson-Crick base pairing. Furthermore, a 2'-OMe modification can affect how the oligonucleotide interacts with transfection reagents, proteins or any other molecules in the cell. The effects of these modifications can be determined by empirical testing.

In some embodiments, the crRNA includes one or more phosphorothioate modifications. In some embodiments, the crRNA includes one or more locked nucleic acids for the purpose of enhancing base pairing and/or increasing nuclease resistance.

A summary of these chemical modifications can be found, *e.g.*, in Kelley *et al.*, "Versatility of chemically synthesized guide RNAs for CRISPR-Cas9 genome editing," *J. Biotechnol.* 233:74-83, 2016; WO 2016205764; and U.S. Pat. No. 8,795,965 B2; each which is incorporated by reference in its entirety.

The sequences and the lengths of the RNA guides (*e.g.*, crRNAs) described herein can be optimized. In some embodiments, the optimized length of an RNA guide can be determined by identifying the processed form of crRNA (*i.e.*, a mature crRNA), or by empirical length studies for crRNA tetraloops.

The crRNAs can also include one or more aptamer sequences. Aptamers are oligonucleotide or peptide molecules have a specific three-dimensional structure and can bind to

a specific target molecule. The aptamers can be specific to gene effectors, gene activators, or gene repressors. In some embodiments, the aptamers can be specific to a protein, which in turn is specific to and recruits and/or binds to specific gene effectors, gene activators, or gene repressors. The effectors, activators, or repressors can be present in the form of fusion proteins. In some embodiments, the guide RNA has two or more aptamer sequences that are specific to the same adaptor proteins. In some embodiments, the two or more aptamer sequences are specific to different adaptor proteins. The adaptor proteins can include, *e.g.*, MS2, PP7, Qβ, F2, GA, fr, JP501, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ϕkCb5, ϕkCb8r, ϕkCb12r, ϕkCb23r, 7s, and PRR1. Accordingly, in some embodiments, the aptamer is selected from binding proteins specifically binding any one of the adaptor proteins as described herein. In some embodiments, the aptamer sequence is a MS2 binding loop (5'-ggcccAACAUGAGGAUCACCCAUGUCUGCAGgggcc-3'). In some embodiments, the aptamer sequence is a QBeta binding loop (5'-ggcccAUGCUGUCUAAGACAGCAUgggcc-3'). In some embodiments, the aptamer sequence is a PP7 binding loop (5'-ggcccUAAGGGUUUAUAUGGAAACCCUUAgggcc-3'. A detailed description of aptamers can be found, *e.g.*, in Nowak *et al.*, "Guide RNA engineering for versatile Cas9 functionality," *Nucl. Acid. Res.*, 44(20):9555-9564, 2016; and WO 2016205764, which are incorporated herein by reference in their entirety.

In certain embodiments, the methods make use of chemically modified guide RNAs. Examples of guide RNA chemical modifications include, without limitation, incorporation of 2'-O-methyl (M), 2'-O-methyl 3'-phosphorothioate (MS), or 2'-O-methyl 3'-thioPACE (MSP) at one or more terminal nucleotides. Such chemically modified guide RNAs can comprise increased stability and increased activity as compared to unmodified guide RNAs, though on-target vs. off- target specificity is not predictable. See, Hendel, *Nat Biotechnol.* 33(9):985-9, 2015, incorporated by reference). Chemically modified guide RNAs may further include, without limitation, RNAs with phosphorothioate linkages and locked nucleic acid (LNA) nucleotides comprising a methylene bridge between the 2' and 4' carbons of the ribose ring.

The invention also encompasses methods for delivering multiple nucleic acid components, wherein each nucleic acid component is specific for a different target locus of interest thereby modifying multiple target loci of interest. The nucleic acid component of the complex may comprise one or more protein-binding RNA aptamers. The one or more aptamers may be

capable of binding a bacteriophage coat protein. The bacteriophage coat protein may be selected from the group comprising Qβ, F2, GA, fr, JP501, MS2, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ϕCb5, ϕCb8r, ϕCb12r, ϕCb23r, 7s and PRR1. In certain embodiments, the bacteriophage coat protein is MS2.

5. Target RNA

The target RNA can be any RNA molecule of interest, including naturally-occurring and engineered RNA molecules. The target RNA can be an mRNA, a tRNA, a ribosomal RNA (rRNA), a microRNA (miRNA), an interfering RNA (siRNA), a ribozyme, a riboswitch, a satellite RNA, a microswitch, a microzyme, or a viral RNA.

In some embodiments, the target nucleic acid is associated with a condition or disease (*e.g.*, an infectious disease or a cancer).

Thus, in some embodiments, the systems described herein can be used to treat a condition or disease by targeting these nucleic acids. For instance, the target nucleic acid associated with a condition or disease may be an RNA molecule that is overexpressed in a diseased cell (*e.g.*, a cancer or tumor cell). The target nucleic acid may also be a toxic RNA and/or a mutated RNA (*e.g.*, an mRNA molecule having a splicing defect or a mutation). The target nucleic acid may also be an RNA that is specific for a particular microorganism (*e.g.*, a pathogenic bacteria).

6. Complex and Cell

One aspect of the invention provides a CRISPR/Cas13e or CRISPR/Cas13f complex comprising (1) any of the Cas13e/Cas13f effector proteins, homologs, orthologs, fusions, derivative, conjugates, or functional fragments thereof as described herein, and (2) any of the guide RNA described herein, each including a spacer sequence designed to be at least partially complementary to a target RNA, and a DR sequence compatible with the Cas13e/Cas13f effector proteins, homologs, orthologs, fusions, derivatives, conjugates, or functional fragments thereof.

In certain embodiments, the complex further comprises the target RNA bound by the guide RNA.

In certain embodiments, the complex is not naturally existing / occurring. For example,

at least one of the components of the complex is not naturally existing / occurring. In certain embodiments, the Cas13e/Cas13f effector protein, homolog, ortholog, fusion, derivative, conjugate, or functional fragment thereof is not naturally occurring / existing due to, for example, the existence of at least one amino acid mutation (deletion, insertion, and/or substitution) as compared to a wild-type protein. In certain embodiments, the DR sequence is not naturally occurring / existing, *i.e.*, not any one of SEQ ID NOs: 8-14, due to, for example, addition, deletion, and/or substitution of at least one nucleotide base in the wild-type sequence. In certain embodiments, the spacer sequence is not naturally occurring, in that it is not present or encoded by any spacer sequences present in the wild-type CRISPR locus of a prokaryote in which the subject Cas13e or Cas13f exists. The spacer sequence may be not naturally existing when it is not 100% complementary to a naturally-occurring bacterialphage nucleic acid.

In a related aspect, the invention also provides a cell comprising any of the complex of the invention.

In certain embodiments, the cell is a prokaryote.

In certain embodiments, the cell is a eukaryote. When the cell is a eukaryote, the complex in the eukaryotic cell can be a naturally existing Cas13e/Cas13f complex in a prokaryote from which the Cas13e/Cas13f is isolated.

7. Methods of Using CRISPR Systems

The CRISPR systems described herein have a wide variety of utilities including modifying (*e.g.*, deleting, inserting, translocating, inactivating, or activating) a target polynucleotide or nucleic acid in a multiplicity of cell types. The CRISPR systems have a broad spectrum of applications in, *e.g.*, DNA/RNA detection (*e.g.*, specific high sensitivity enzymatic reporter unlocking (SHERLOCK)), tracking and labeling of nucleic acids, enrichment assays (extracting desired sequence from background), controlling interfering RNA or miRNA, detecting circulating tumor DNA, preparing next generation library, drug screening, disease diagnosis and prognosis, and treating various genetic disorders.

DNA/RNA Detection

In one aspect, the CRISPR systems described herein can be used in DNA or RNA

detection. As shown in the examples, the Cas13e and Cas13f proteins of the invention exhibit non-specific / collateral RNase activity upon activation of its guide RNA-dependent specific RNase activity when the spacer sequence is about 30 nucleotides. Thus the CRISPR-associated proteins of the invention can be reprogrammed with CRISPR RNAs (crRNAs) to provide a platform for specific RNA sensing. By choosing specific spacer sequence length, and upon recognition of its RNA target, activated CRISPR-associated proteins engage in "collateral" cleavage of nearby non-targeted RNAs. This crRNA-programmed collateral cleavage activity allows the CRISPR systems to detect the presence of a specific RNA by triggering programmed cell death or by nonspecific degradation of labeled RNA.

The SHERLOCK method (Specific High Sensitivity Enzymatic Reporter UnLOCKing) provides an *in vitro* nucleic acid detection platform with attomolar sensitivity based on nucleic acid amplification and collateral cleavage of a reporter RNA, allowing for real-time detection of the target. To achieve signal detection, the detection can be combined with different isothermal amplification steps. For example, recombinase polymerase amplification (RPA) can be coupled with T7 transcription to convert amplified DNA to RNA for subsequent detection. The combination of amplification by RPA, T7 RNA polymerase transcription of amplified DNA to RNA, and detection of target RNA by collateral RNA cleavage-mediated release of reporter signal is referred as SHERLOCK. Methods of using CRISPR in SHERLOCK are described in detail, *e.g.*, in Gootenberg, *et al.* "Nucleic acid detection with CRISPR-Cas13a/C2c2," Science, 2017 Apr. 28; 356(6336):438-442, which is incorporated herein by reference in its entirety.

The CRISPR-associated proteins can be used in Northern blot assays, which use electrophoresis to separate RNA samples by size. The CRISPR-associated proteins can be used to specifically bind and detect the target RNA sequence. The CRISPR-associated proteins can also be fused to a fluorescent protein (*e.g.*, GFP) and used to track RNA localization in living cells. More particularly, the CRISPR-associated proteins can be inactivated in that they no longer cleave RNAs as described above. Thus, CRISPR-associated proteins can be used to determine the localization of the RNA or specific splice variants, the level of mRNA transcripts, up- or down-regulation of transcripts and disease-specific diagnosis. The CRISPR-associated proteins can be used for visualization of RNA in (living) cells using, for example, fluorescent microscopy or flow cytometry, such as fluorescence-activated cell sorting (FACS), which allows for high-throughput screening of cells and recovery of living cells following cell sorting. A

detailed description regarding how to detect DNA and RNA can be found, *e.g.*, in International Publication No. WO 2017/070605, which is incorporated herein by reference in its entirety.

In some embodiments, the CRISPR systems described herein can be used in multiplexed error-robust fluorescence in situ hybridization (MERFISH). These methods are described in, *e.g.*, Chen *et al.*, "Spatially resolved, highly multiplexed RNA profiling in single cells," Science, 2015 Apr. 24; 348(6233):aaa6090, which is incorporated herein by reference herein in its entirety.

In some embodiments, the CRISPR systems described herein can be used to detect a target RNA in a sample (*e.g.*, a clinical sample, a cell, or a cell lysate). The collateral RNase activity of the Type VI-E and/or VI-F CRISPR-Cas effector proteins described herein is activated when the effector proteins bind to a target nucleic acid when the spacer sequence is of a specific chosen length (such as about 30 nucleotides). Upon binding to the target RNA of interest, the effector protein cleaves a labeled detector RNA to generate a signal (*e.g.*, an increased signal or a decreased signal) thereby allowing for the qualitative and quantitative detection of the target RNA in the sample. The specific detection and quantification of RNA in the sample allows for a multitude of applications including diagnostics. In some embodiments, the methods include contacting a sample with: i) an RNA guide (*e.g.*, crRNA) and/or a nucleic acid encoding the RNA guide, wherein the RNA guide consists of a direct repeat sequence and a spacer sequence capable of hybridizing to the target RNA; (ii) a Type VI-E or VI-F CRISPR-Cas effector protein (Cas13e or Cas13f) and/or a nucleic acid encoding the effector protein; and (iii) a labeled detector RNA; wherein the effector protein associates with the RNA guide to form a complex; wherein the RNA guide hybridizes to the target RNA; and wherein upon binding of the complex to the target RNA, the effector protein exhibits collateral RNase activity and cleaves the labeled detector RNA; and b) measuring a detectable signal produced by cleavage of the labeled detector RNA, wherein said measuring provides for detection of the single-stranded target RNA in the sample. In some embodiments, the methods further comprise comparing the detectable signal with a reference signal and determining the amount of target RNA in the sample. In some embodiments, the measuring is performed using gold nanoparticle detection, fluorescence polarization, colloid phase transition/dispersion, electrochemical detection, and semiconductor based-sensing. In some embodiments, the labeled detector RNA includes a fluorescence-emitting dye pair, a fluorescence resonance energy transfer (FRET) pair, or a quencher/fluor pair. In some embodiments, upon cleavage of the labeled detector RNA by the effector protein, an amount of

detectable signal produced by the labeled detector RNA is decreased or increased. In some embodiments, the labeled detector RNA produces a first detectable signal prior to cleavage by the effector protein and a second detectable signal after cleavage by the effector protein. In some embodiments, a detectable signal is produced when the labeled detector RNA is cleaved by the effector protein. In some embodiments, the labeled detector RNA comprises a modified nucleobase, a modified sugar moiety, a modified nucleic acid linkage, or a combination thereof. In some embodiments, the methods include the multi-channel detection of multiple independent target RNAs in a sample (*e.g.*, two, three, four, five, six, seven, eight, nine, ten, fifteen, twenty, thirty, forty, or more target RNAs) by using multiple Type VI-E and/or VI-F CRISPR-Cas (Cas13e and/or Cas13f) systems, each including a distinct orthologous effector protein and corresponding RNA guides, allowing for the differentiation of multiple target RNAs in the sample. In some embodiments, the methods include the multi-channel detection of multiple independent target RNAs in a sample, with the use of multiple instances of Type VI-E and/or VI-F CRISPR-Cas systems, each containing an orthologous effector protein with differentiable collateral RNase substrates. Methods of detecting an RNA in a sample using CRISPR-associated proteins are described, for example, in U.S. Patent Publication No. 2017/0362644, the entire contents of which are incorporated herein by reference.

Tracking and Labeling of Nucleic Acids

Cellular processes depend on a network of molecular interactions among proteins, RNAs, and DNAs. Accurate detection of protein-DNA and protein-RNA interactions is key to understanding such processes. *In vitro* proximity labeling techniques employ an affinity tag combined with, a reporter group, *e.g.*, a photoactivatable group, to label polypeptides and RNAs in the vicinity of a protein or RNA of interest *in vitro*. After UV irradiation, the photoactivatable groups react with proteins and other molecules that are in close proximity to the tagged molecules, thereby labelling them. Labelled interacting molecules can subsequently be recovered and identified. The CRISPR-associated proteins can for instance be used to target probes to selected RNA sequences. These applications can also be applied in animal models for *in vivo* imaging of diseases or difficult-to culture cell types. The methods of tracking and labeling of nucleic acids are described, *e.g.*, in U.S. Pat. No. 8,795,965, WO 2016205764, and WO 2017070605; each of which is incorporated herein by reference herein in its entirety.

RNA Isolation, Purification, Enrichment, and/or Depletion

The CRISPR systems (*e.g.*, CRISPR-associated proteins) described herein can be used to isolate and/or purify the RNA. The CRISPR-associated proteins can be fused to an affinity tag that can be used to isolate and/or purify the RNA-CRISPR-associated protein complex. These applications are useful, *e.g.*, for the analysis of gene expression profiles in cells.

In some embodiments, the CRISPR-associated proteins can be used to target a specific noncoding RNA (ncRNA) thereby blocking its activity. In some embodiments, the CRISPR-associated proteins can be used to specifically enrich a particular RNA (including but not limited to increasing stability, *etc.*), or alternatively, to specifically deplete a particular RNA (*e.g.*, particular splice variants, isoforms, *etc.*).

These methods are described, *e.g.*, in U.S. Pat. No. 8,795,965, WO 2016205764, and WO 2017070605; each of which is incorporated herein by reference herein in its entirety.

High-Throughput Screening

The CRISPR systems described herein can be used for preparing next generation sequencing (NGS) libraries. For example, to create a cost-effective NGS library, the CRISPR systems can be used to disrupt the coding sequence of a target gene, and the CRISPR-associated protein transfected clones can be screened simultaneously by next-generation sequencing (*e.g.*, on the Ion Torrent PGM system). A detailed description regarding how to prepare NGS libraries can be found, *e.g.*, in Bell *et al.*, "A high-throughput screening strategy for detecting CRISPR-Cas9 induced mutations using next-generation sequencing," BMC Genomics, 15.1 (2014): 1002, which is incorporated herein by reference in its entirety.

Engineered Microorganisms

Microorganisms (*e.g.*, *E. coli*, yeast, and microalgae) are widely used for synthetic biology. The development of synthetic biology has a wide utility, including various clinical applications. For example, the programmable CRISPR systems can be used to split proteins of toxic domains for targeted cell death, *e.g.*, using cancer-linked RNA as target transcript. Further, pathways involving protein-protein interactions can be influenced in synthetic biological systems with, *e.g.*, fusion complexes with the appropriate effectors such as kinases or enzymes.

In some embodiments, crRNAs that target phage sequences can be introduced into the microorganism. Thus, the disclosure also provides methods of vaccinating a microorganism (*e.g.*, a production strain) against phage infection.

In some embodiments, the CRISPR systems provided herein can be used to engineer microorganisms, *e.g.*, to improve yield or improve fermentation efficiency. For example, the CRISPR systems described herein can be used to engineer microorganisms, such as yeast, to generate biofuel or biopolymers from fermentable sugars, or to degrade plant-derived lignocellulose derived from agricultural waste as a source of fermentable sugars. More particularly, the methods described herein can be used to modify the expression of endogenous genes required for biofuel production and/or to modify endogenous genes, which may interfere with the biofuel synthesis. These methods of engineering microorganisms are described *e.g.*, in Verwaal *et al.*, "CRISPR/Cpf1 enables fast and simple genome editing of Saccharomyces cerevisiae," Yeast doi: 10.1002/yea.3278, 2017; and Hlavova *et al.*, "Improving microalgae for biotechnology-from genetics to synthetic biology," Biotechnol. Adv., 33:1194-203, 2015, both of which are incorporated herein by reference in the entirety.

In some embodiments, the CRISPR systems provided herein can be used to induce death or dormancy of a cell (*e.g.*, a microorganism such as an engineered microorganism). These methods can be used to induce dormancy or death of a multitude of cell types including prokaryotic and eukaryotic cells, including, but not limited to mammalian cells (*e.g.*, cancer cells, or tissue culture cells), protozoans, fungal cells, cells infected with a virus, cells infected with an intracellular bacteria, cells infected with an intracellular protozoan, cells infected with a prion, bacteria (*e.g.*, pathogenic and non-pathogenic bacteria), protozoans, and unicellular and multicellular parasites. For instance, in the field of synthetic biology it is highly desirable to have mechanisms of controlling engineered microorganisms (*e.g.*, bacteria) in order to prevent their propagation or dissemination. The systems described herein can be used as "kill-switches" to regulate and/or prevent the propagation or dissemination of an engineered microorganism. Further, there is a need in the art for alternatives to current antibiotic treatments. The systems described herein can also be used in applications where it is desirable to kill or control a specific microbial population (*e.g.*, a bacterial population). For example, the systems described herein may include an RNA guide (*e.g.*, a crRNA) that targets a nucleic acid (*e.g.*, an RNA) that is genus-, species-, or strain-specific, and can be delivered to the cell. Upon complexing and binding to the target nucleic acid, the collateral RNase activity of the Type VI-E and/or VI-F CRISPR-Cas effector proteins is activated leading to the cleavage of non-target RNA within the microorganisms, ultimately resulting in dormancy or death. In some embodiments, the methods

comprise contacting the cell with a system described herein including a Type VI-E and/or VI-F CRISPR-Cas effector proteins or a nucleic acid encoding the effector protein, and a RNA guide (*e.g.*, a crRNA) or a nucleic acid encoding the RNA guide, wherein the spacer sequence is complementary to at least 15 nucleotides (*e.g.*, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50 or more nucleotides) of a target nucleic acid (*e.g.*, a genus-, strain-, or species-specific RNA guide). Without wishing to be bound by any particular theory, the cleavage of non-target RNA by the Type VI-E and/or VI-F CRISPR-Cas effector proteins may induce programmed cell death, cell toxicity, apoptosis, necrosis, necroptosis, cell death, cell cycle arrest, cell anergy, a reduction of cell growth, or a reduction in cell proliferation. For example, in bacteria, the cleavage of non-target RNA by the Type VI-E and/or VI-F CRISPR-Cas effector proteins may be bacteriostatic or bactericidal.

Application in Plants

The CRISPR systems described herein have a wide variety of utility in plants. In some embodiments, the CRISPR systems can be used to engineer genomes of plants (*e.g.*, improving production, making products with desired post-translational modifications, or introducing genes for producing industrial products). In some embodiments, the CRISPR systems can be used to introduce a desired trait to a plant (*e.g.*, with or without heritable modifications to the genome), or regulate expression of endogenous genes in plant cells or whole plants.

In some embodiments, the CRISPR systems can be used to identify, edit, and/or silence genes encoding specific proteins, *e.g.*, allergenic proteins (*e.g.*, allergenic proteins in peanuts, soybeans, lentils, peas, green beans, and mung beans). A detailed description regarding how to identify, edit, and/or silence genes encoding proteins is described, *e.g.*, in Nicolaou *et al.*, "Molecular diagnosis of peanut and legume allergy," *Curr. Opin. Allergy Clin. Immunol.* 11(3):222-8, 2011, and WO 2016205764 A1; both of which are incorporated herein by reference in the entirety.

Gene Drives

Gene drive is the phenomenon in which the inheritance of a particular gene or set of genes is favorably biased. The CRISPR systems described herein can be used to build gene drives. For example, the CRISPR systems can be designed to target and disrupt a particular allele of a gene, causing the cell to copy the second allele to fix the sequence. Because of the

copying, the first allele will be converted to the second allele, increasing the chance of the second allele being transmitted to the offspring. A detailed method regarding how to use the CRISPR systems described herein to build gene drives is described, *e.g.*, in Hammond *et al.*, "A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae," *Nat. Biotechnol.* 34(1):78-83, 2016, which is incorporated herein by reference in its entirety.

Pooled-Screening

As described herein, pooled CRISPR screening is a powerful tool for identifying genes involved in biological mechanisms such as cell proliferation, drug resistance, and viral infection. Cells are transduced in bulk with a library of guide RNA (gRNA)-encoding vectors described herein, and the distribution of gRNAs is measured before and after applying a selective challenge. Pooled CRISPR screens work well for mechanisms that affect cell survival and proliferation, and they can be extended to measure the activity of individual genes (*e.g.*, by using engineered reporter cell lines). Arrayed CRISPR screens, in which only one gene is targeted at a time, make it possible to use RNA-seq as the readout. In some embodiments, the CRISPR systems as described herein can be used in single-cell CRISPR screens. A detailed description regarding pooled CRISPR screenings can be found, *e.g.*, in Datlinger *et al.*, "Pooled CRISPR screening with single-cell transcriptome read-out," *Nat. Methods.* 14(3):297-301, 2017, which is incorporated herein by reference in its entirety.

Saturation Mutagenesis (Bashing)

The CRISPR systems described herein can be used for *in situ* saturating mutagenesis. In some embodiments, a pooled guide RNA library can be used to perform *in situ* saturating mutagenesis for particular genes or regulatory elements. Such methods can reveal critical minimal features and discrete vulnerabilities of these genes or regulatory elements (*e.g.*, enhancers). These methods are described, *e.g.*, in Canver *et al.*, "BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis," *Nature* 527(7577):192-7, 2015, which is incorporated herein by reference in its entirety.

RNA-Related Applications

The CRISPR systems described herein can have various RNA-related applications, *e.g.*, modulating gene expression, degrading a RNA molecule, inhibiting RNA expression, screening

RNA or RNA products, determining functions of lincRNA or non-coding RNA, inducing cell dormancy, inducing cell cycle arrest, reducing cell growth and/or cell proliferation, inducing cell anergy, inducing cell apoptosis, inducing cell necrosis, inducing cell death, and/or inducing programmed cell death. A detailed description of these applications can be found, *e.g.*, in WO 2016/205764 A1, which is incorporated herein by reference in its entirety. In different embodiments, the methods described herein can be performed *in vitro*, *in vivo*, or *ex vivo*.

For example, the CRISPR systems described herein can be administered to a subject having a disease or disorder to target and induce cell death in a cell in a diseased state (*e.g.*, cancer cells or cells infected with an infectious agent). For instance, in some embodiments, the CRISPR systems described herein can be used to target and induce cell death in a cancer cell, wherein the cancer cell is from a subject having a Wilms' tumor, Ewing sarcoma, a neuroendocrine tumor, a glioblastoma, a neuroblastoma, a melanoma, skin cancer, breast cancer, colon cancer, rectal cancer, prostate cancer, liver cancer, renal cancer, pancreatic cancer, lung cancer, biliary cancer, cervical cancer, endometrial cancer, esophageal cancer, gastric cancer, head and neck cancer, medullary thyroid carcinoma, ovarian cancer, glioma, lymphoma, leukemia, myeloma, acute lymphoblastic leukemia, acute myelogenous leukemia, chronic lymphocytic leukemia, chronic myelogenous leukemia, Hodgkin's lymphoma, non-Hodgkin's lymphoma, or urinary bladder cancer.

Modulating Gene Expression

The CRISPR systems described herein can be used to modulate gene expression. The CRISPR systems can be used, together with suitable guide RNAs, to target gene expression, via control of RNA processing. The control of RNA processing can include, *e.g.*, RNA processing reactions such as RNA splicing (*e.g.*, alternative splicing), viral replication, and tRNA biosynthesis. The RNA targeting proteins in combination with suitable guide RNAs can also be used to control RNA activation (RNAa). RNA activation is a small RNA-guided and Argonaute (Ago)-dependent gene regulation phenomenon in which promoter-targeted short double-stranded RNAs (dsRNAs) induce target gene expression at the transcriptional/epigenetic level. RNAa leads to the promotion of gene expression, so control of gene expression may be achieved that way through disruption or reduction of RNAa. In some embodiments, the methods include the use of the RNA targeting CRISPR as substitutes for *e.g.*, interfering ribonucleic acids (such as siRNAs, shRNAs, or dsRNAs). The methods of modulating gene expression are described, *e.g.*,

in WO 2016205764, which is incorporated herein by reference in its entirety.

Controlling RNA Interference

Control over interfering RNAs or microRNAs (miRNA) can help reduce off-target effects by reducing the longevity of the interfering RNAs or miRNAs *in vivo* or *in vitro*. In some embodiments, the target RNAs can include interfering RNAs, *i.e.*, RNAs involved in the RNA interference pathway, such as small hairpin RNAs (shRNAs), small interfering (siRNAs), *etc.* In some embodiments, the target RNAs include, *e.g.*, miRNAs or double stranded RNAs (dsRNA).

In some embodiments, if the RNA targeting protein and suitable guide RNAs are selectively expressed (for example spatially or temporally under the control of a regulated promoter, for example a tissue- or cell cycle-specific promoter and/or enhancer), this can be used to protect the cells or systems (*in vivo* or *in vitro*) from RNA interference (RNAi) in those cells. This may be useful in neighboring tissues or cells where RNAi is not required or for the purposes of comparison of the cells or tissues where the CRISPR-associated proteins and suitable crRNAs are and are not expressed (*i.e.*, where the RNAi is not controlled and where it is, respectively). The RNA targeting proteins can be used to control or bind to molecules comprising or consisting of RNAs, such as ribozymes, ribosomes, or riboswitches. In some embodiments, the guide RNAs can recruit the RNA targeting proteins to these molecules so that the RNA targeting proteins are able to bind to them. These methods are described, *e.g.*, in WO 2016205764 and WO 2017070605, both of which are incorporated herein by reference in the entirety.

Modifying Riboswitches and Controlling Metabolic Regulations

Riboswitches are regulatory segments of messenger RNAs that bind small molecules and in turn regulate gene expression. This mechanism allows the cell to sense the intracellular concentration of these small molecules. A specific riboswitch typically regulates its adjacent gene by altering the transcription, the translation or the splicing of this gene. Thus, in some embodiments, the riboswitch activity can be controlled by the use of the RNA targeting proteins in combination with suitable guide RNAs to target the riboswitches. This may be achieved through cleavage of, or binding to, the riboswitch. Methods of using CRISPR systems to control riboswitches are described, *e.g.*, in WO 2016205764 and WO 2017070605, both of which are incorporated herein by reference in their entireties.

RNA Modification

In some embodiments, the CRISPR-associated proteins described herein can be fused to a base-editing domain, such as ADAR1, ADAR2, APOBEC, or activation-induced cytidine deaminase (AID), and can be used to modify an RNA sequence (*e.g.*, an mRNA). In some embodiments, the CRISPR-associated protein includes one or more mutations (*e.g.*, in a catalytic domain), which renders the CRISPR-associated protein incapable of cleaving RNA.

In some embodiments, the CRISPR-associated proteins can be used with an RNA-binding fusion polypeptide comprising a base-editing domain (*e.g.*, ADAR1, ADAR2, APOBEC, or AID) fused to an RNA-binding domain, such as MS2 (also known as MS2 coat protein), Qbeta (also known as Qbeta coat protein), or PP7 (also known as PP7 coat protein). The amino acid sequences of the RNA-binding domains MS2, Qbeta, and PP7 are provided below:

MS2 (MS2 coat protein)

MASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQKRKYTIKVEVPKVAT
QTVGGVELPVAAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY

Qbeta (Qbeta coat protein)

MAKLETVTLGNIGKDGKQTLVLNPRGVNPTNGVASLSQAGAVPALEKRVTVSVSQPSRNRKNYKVQVKIQ
NPTACTANGSCDPSVTRQAYADVTFSFTQYSTDEERAFVRTELAALLASPLLIDAIDQLNPAY

PP7 (PP7 coat protein)

MSKTIVLSVGEATRTLTEIQSTADRQIFEEKVGPLVGRLRLTASLRQNGAKTAYRVNLKLDQADVVDCST
SVCGELPKVRYTQVWSHDVTIVANSTEASRKSLYDLTKSLVVQATSEDLVVNLVPLGR

In some embodiments, the RNA binding domain can bind to a specific sequence (*e.g.*, an aptamer sequence) or secondary structure motifs on a crRNA of the system described herein (*e.g.*, when the crRNA is in an effector-crRNA complex), thereby recruiting the RNA binding fusion polypeptide (which has a base-editing domain) to the effector complex. For example, in some embodiments, the CRISPR system includes a CRISPR associated protein, a crRNA having an aptamer sequence (*e.g.*, an MS2 binding loop, a QBeta binding loop, or a PP7 binding loop), and a RNA-binding fusion polypeptide having a base-editing domain fused to an RNA-binding domain that specifically binds to the aptamer sequence. In this system, the CRISPR-associated protein forms a complex with the crRNA having the aptamer sequence. Further the RNA-binding fusion polypeptide binds to the crRNA (via the aptamer sequence) thereby forming a tripartite complex that can modify a target RNA.

Methods of using CRISPR systems for base editing are described, *e.g.*, in International Publication No. WO 2017/219027, which is incorporated herein by reference in its entirety, and in particular with respect to its discussion of RNA modification.

RNA Splicing

In some embodiments, an inactivated CRISPR-associated protein described herein (*e.g.*, a CRISPR associated protein having one or more mutations in a catalytic domain) can be used to target and bind to specific splicing sites on RNA transcripts. Binding of the inactivated CRISPR-associated protein to the RNA may sterically inhibit interaction of the spliceosome with the transcript, enabling alteration in the frequency of generation of specific transcript isoforms. Such method can be used to treat disease through exon skipping such that an exon having a mutation may be skipped in a mature protein. Methods of using CRISPR systems to alter splicing are described, *e.g.*, in International Publication No. WO 2017/219027, which is incorporated herein by reference in its entirety, and in particular with respect to its discussion of RNA splicing.

Therapeutic Applications

The CRISPR systems described herein can have various therapeutic applications. Such applications may be based on one or more of the abilities below, both *in vitro* and *in vivo*, of the subject CRISPR/Cas13e or Cas13f systems: induce cellular senescence, induce cell cycle arrest, inhibit cell growth and/or proliferation, induce apoptosis, induce necrosis, *etc.*

In some embodiments, the new CRISPR systems can be used to treat various diseases and disorders, *e.g.*, genetic disorders (*e.g.*, monogenetic diseases), diseases that can be treated by nuclease activity (*e.g.*, Pcsk9 targeting, Duchenne Muscular Dystrophy (DMD), BCL11a targeting), and various cancers, *etc.*

In some embodiments, the CRISPR systems described herein can be used to edit a target nucleic acid to modify the target nucleic acid (*e.g.*, by inserting, deleting, or mutating one or more nucleic acid residues). For example, in some embodiments the CRISPR systems described herein comprise an exogenous donor template nucleic acid (*e.g.*, a DNA molecule or a RNA molecule), which comprises a desirable nucleic acid sequence. Upon resolution of a cleavage event induced with the CRISPR system described herein, the molecular machinery of the cell will utilize the exogenous donor template nucleic acid in repairing and/or resolving the cleavage

event. Alternatively, the molecular machinery of the cell can utilize an endogenous template in repairing and/or resolving the cleavage event. In some embodiments, the CRISPR systems described herein may be used to alter a target nucleic acid resulting in an insertion, a deletion, and/or a point mutation). In some embodiments, the insertion is a scarless insertion (*i.e.*, the insertion of an intended nucleic acid sequence into a target nucleic acid resulting in no additional unintended nucleic acid sequence upon resolution of the cleavage event). Donor template nucleic acids may be double stranded or single stranded nucleic acid molecules (*e.g.*, DNA or RNA). Methods of designing exogenous donor template nucleic acids are described, for example, in International Publication No. WO 2016/094874 A1, the entire contents of which are expressly incorporated herein by reference.

In one aspect, the CRISPR systems described herein can be used for treating a disease caused by overexpression of RNAs, toxic RNAs, and/or mutated RNAs (*e.g.*, splicing defects or truncations). For example, expression of toxic RNAs may be associated with the formation of nuclear inclusions and late-onset degenerative changes in brain, heart, or skeletal muscle. In some embodiments, the disorder is myotonic dystrophy. In myotonic dystrophy, the main pathogenic effect of the toxic RNAs is to sequester binding proteins and compromise the regulation of alternative splicing (see, *e.g.*, Osborne *et al.*, "RNA-dominant diseases," Hum. Mol. Genet., 2009 Apr. 15; 18(8):1471-81). Myotonic dystrophy (dystrophia myotonica (DM)) is of particular interest to geneticists because it produces an extremely wide range of clinical features. The classical form of DM, which is now called DM type 1 (DM1), is caused by an expansion of CTG repeats in the 3'-untranslated region (UTR) of DMPK, a gene encoding a cytosolic protein kinase. The CRISPR systems as described herein can target overexpressed RNA or toxic RNA, *e.g.*, the DMPK gene or any of the mis-regulated alternative splicing in DM1 skeletal muscle, heart, or brain.

The CRISPR systems described herein can also target trans-acting mutations affecting RNA-dependent functions that cause various diseases such as, *e.g.*, Prader Willi syndrome, Spinal muscular atrophy (SMA), and Dyskeratosis congenita. A list of diseases that can be treated using the CRISPR systems described herein is summarized in Cooper *et al.*, "RNA and disease," Cell, 136.4 (2009): 777-793, and WO 2016/205764 A1, both of which are incorporated herein by reference in the entirety. Those of skill in this field will understand how to use the new CRISPR systems to treat these diseases.

The CRISPR systems described herein can also be used in the treatment of various tauopathies, including, *e.g.*, primary and secondary tauopathies, such as primary age-related tauopathy (PART)/Neurofibrillary tangle (NFT)-predominant senile dementia (with NFTs similar to those seen in Alzheimer Disease (AD), but without plaques), dementia pugilistica (chronic traumatic encephalopathy), and progressive supranuclear palsy. A useful list of tauopathies and methods of treating these diseases are described, *e.g.*, in WO 2016205764, which is incorporated herein by reference in its entirety.

The CRISPR systems described herein can also be used to target mutations disrupting the cis-acting splicing codes that can cause splicing defects and diseases. These diseases include, *e.g.*, motor neuron degenerative disease that results from deletion of the SMN1 gene (*e.g.*, spinal muscular atrophy), Duchenne Muscular Dystrophy (DMD), frontotemporal dementia, and Parkinsonism linked to chromosome 17 (FTDP-17), and cystic fibrosis.

The CRISPR systems described herein can further be used for antiviral activity, in particular against RNA viruses. The CRISPR-associated proteins can target the viral RNAs using suitable guide RNAs selected to target viral RNA sequences.

The CRISPR systems described herein can also be used to treat a cancer in a subject (*e.g.*, a human subject). For example, the CRISPR-associated proteins described herein can be programmed with crRNA targeting a RNA molecule that is aberrant (*e.g.*, comprises a point mutation or are alternatively-spliced) and found in cancer cells to induce cell death in the cancer cells (*e.g.*, via apoptosis).

The CRISPR systems described herein can also be used to treat an autoimmune disease or disorder in a subject (*e.g.*, a human subject). For example, the CRISPR-associated proteins described herein can be programmed with crRNA targeting a RNA molecule that is aberrant (*e.g.*, comprises a point mutation or are alternatively-spliced) and found in cells responsible for causing the autoimmune disease or disorder.

Further, the CRISPR systems described herein can also be used to treat an infectious disease in a subject. For example, the CRISPR-associated proteins described herein can be programmed with crRNA targeting a RNA molecule expressed by an infectious agent (*e.g.*, a bacteria, a virus, a parasite or a protozoan) in order to target and induce cell death in the infectious agent cell. The CRISPR systems may also be used to treat diseases where an

intracellular infectious agent infects the cells of a host subject. By programming the CRISPR-associated protein to target a RNA molecule encoded by an infectious agent gene, cells infected with the infectious agent can be targeted and cell death induced.

Furthermore, *in vitro* RNA sensing assays can be used to detect specific RNA substrates. The CRISPR-associated proteins can be used for RNA-based sensing in living cells. Examples of applications are diagnostics by sensing of, for examples, disease-specific RNAs.

A detailed description of therapeutic applications of the CRISPR systems described herein can be found, *e.g.*, in U.S. Pat. No. 8,795,965, EP 3009511, WO 2016205764, and WO 2017070605; each of which is incorporated herein by reference in its entirety.

Cells and Progenies Thereof

In certain embodiments, the methods of the invention can be used to introduce the CRISPR systems described herein into a cell, and cause the cell and/or its progeny to alter the production of one or more cellular produces, such as antibody, starch, ethanol, or any other desired products. Such cells and progenies thereof are within the scope of the invention.

In certain embodiments, the methods and/or the CRISPR systems described herein lead to modification of the translation and/or transcription of one or more RNA products of the cells. For example, the modification may lead to increased transcription / translation / expression of the RNA product. In other embodiments, the modification may lead to decreased transcription / translation / expression of the RNA product.

In certain embodiments, the cell is a prokaryotic cell.

In certain embodiments, the cell is a eukaryotic cell, such as a mammalian cell, including a human cell (a primary human cell or an established human cell line). In certain embodiments, the cell is a non-human mammalian cell, such as a cell from a non-human primate (*e.g.*, monkey), a cow / bull / cattle, sheep, goat, pig, horse, dog, cat, rodent (such as rabbit, mouse, rat, hamster, *etc*). In certain embodiments, the cell is from fish (such as salmon), bird (such as poultry bird, including chick, duck, goose), reptile, shellfish (*e.g.*, oyster, claim, lobster, shrimp), insect, worm, yeast, *etc*. In certain embodiments, the cell is from a plant, such as monocot or dicot. In certain embodiment, the plant is a food crop such as barley, cassava, cotton, groundnuts or peanuts, maize, millet, oil palm fruit, potatoes, pulses, rapeseed or canola, rice, rye, sorghum, soybeans, sugar cane, sugar beets, sunflower, and wheat. In certain embodiment, the plant is a cereal

(barley, maize, millet, rice, rye, sorghum, and wheat). In certain embodiment, the plant is a tuber (cassava and potatoes). In certain embodiment, the plant is a sugar crop (sugar beets and sugar cane). In certain embodiment, the plant is an oil-bearing crop (soybeans, groundnuts or peanuts, rapeseed or canola, sunflower, and oil palm fruit). In certain embodiment, the plant is a fiber crop (cotton). In certain embodiment, the plant is a tree (such as a peach or a nectarine tree, an apple or pear tree, a nut tree such as almond or walnut or pistachio tree, or a citrus tree, *e.g.*, orange, grapefruit or lemon tree), a grass, a vegetable, a fruit, or an algae. In certain embodiment, the plant is a nightshade plant; a plant of the genus *Brassica*; a plant of the genus *Lactuca*; a plant of the genus *Spinacia*; a plant of the genus *Capsicum*; cotton, tobacco, asparagus, carrot, cabbage, broccoli, cauliflower, tomato, eggplant, pepper, lettuce, spinach, strawberry, blueberry, raspberry, blackberry, grape, coffee, cocoa, *etc*.

A related aspect provides cells or progenies thereof modified by the methods of the invention using the CRISPR systems described herein.

In certain embodiments, the cell is modified *in vitro*, *in vivo*, or ex vivo.

In certain embodiments, the cell is a stem cell.

7. Delivery

Through this disclosure and the knowledge in the art, the CRISPR systems described herein, or any of the components thereof described herein (Cas proteins, derivatives, functional fragments or the various fusions or adducts thereof, and guide RNA / crRNA), nucleic acid molecules thereof, and/or nucleic acid molecules encoding or providing components thereof, can be delivered by various delivery systems such as vectors, *e.g.*, plasmids and viral delivery vectors, using any suitable means in the art. Such methods include (and are not limited to) electroporation, lipofection, microinjection, transfection, sonication, gene gun, *etc*.

In certain embodiments, the CRISPR-associated proteins and/or any of the RNAs (*e.g.*, guide RNAs or crRNAs) and/or accessory proteins can be delivered using suitable vectors, *e.g.*, plasmids or viral vectors, such as adeno-associated viruses (AAV), lentiviruses, adenoviruses, retroviral vectors, and other viral vectors, or combinations thereof. The proteins and one or more crRNAs can be packaged into one or more vectors, *e.g.*, plasmids or viral vectors. For bacterial applications, the nucleic acids encoding any of the components of the CRISPR systems described

herein can be delivered to the bacteria using a phage. Exemplary phages, include, but are not limited to, T4 phage, Mu, λ phage, T5 phage, T7 phage, T3 phage, Φ29, M13, MS2, Qβ, and ΦX174.

In some embodiments, the vectors, *e.g.*, plasmids or viral vectors, are delivered to the tissue of interest by, *e.g.*, intramuscular injection, intravenous administration, transdermal administration, intranasal administration, oral administration, or mucosal administration. Such delivery may be either via a single dose, or multiple doses. One skilled in the art understands that the actual dosage to be delivered herein may vary greatly depending upon a variety of factors, such as the vector choices, the target cells, organisms, tissues, the general conditions of the subject to be treated, the degrees of transformation/modification sought, the administration routes, the administration modes, the types of transformation/modification sought, *etc.*

In certain embodiments, the delivery is via adenoviruses, which can be at a single dose containing at least 1×10^5 particles (also referred to as particle units, pu) of adenoviruses. In some embodiments, the dose preferably is at least about 1×10^6 particles, at least about 1×10^7 particles, at least about 1×10^8 particles, and at least about 1×10^9 particles of the adenoviruses. The delivery methods and the doses are described, *e.g.*, in WO 2016205764 A1 and U.S. Pat. No. 8,454,972 B2, both of which are incorporated herein by reference in the entirety.

In some embodiments, the delivery is via plasmids. The dosage can be a sufficient number of plasmids to elicit a response. In some cases, suitable quantities of plasmid DNA in plasmid compositions can be from about 0.1 to about 2 mg. Plasmids will generally include (i) a promoter; (ii) a sequence encoding a nucleic acid-targeting CRISPR-associated proteins and/or an accessory protein, each operably linked to a promoter (*e.g.*, the same promoter or a different promoter); (iii) a selectable marker; (iv) an origin of replication; and (v) a transcription terminator downstream of and operably linked to (ii). The plasmids can also encode the RNA components of a CRISPR complex, but one or more of these may instead be encoded on different vectors. The frequency of administration is within the ambit of the medical or veterinary practitioner (*e.g.*, physician, veterinarian), or a person skilled in the art.

In another embodiment, the delivery is via liposomes or lipofection formulations and the like, and can be prepared by methods known to those skilled in the art. Such methods are described, for example, in WO 2016205764 and U.S. Pat. Nos. 5,593,972; 5,589,466; and 5,580,859; each of which is incorporated herein by reference in its entirety.

In some embodiments, the delivery is via nanoparticles or exosomes. For example, exosomes have been shown to be particularly useful in delivery RNA.

Further means of introducing one or more components of the new CRISPR systems to the cell is by using cell penetrating peptides (CPP). In some embodiments, a cell penetrating peptide is linked to the CRISPR-associated proteins. In some embodiments, the CRISPR-associated proteins and/or guide RNAs are coupled to one or more CPPs to effectively transport them inside cells (*e.g.*, plant protoplasts). In some embodiments, the CRISPR-associated proteins and/or guide RNA(s) are encoded by one or more circular or non-circular DNA molecules that are coupled to one or more CPPs for cell delivery.

CPPs are short peptides of fewer than 35 amino acids derived either from proteins or from chimeric sequences capable of transporting biomolecules across cell membrane in a receptor independent manner. CPPs can be cationic peptides, peptides having hydrophobic sequences, amphipathic peptides, peptides having proline-rich and anti-microbial sequences, and chimeric or bipartite peptides. Examples of CPPs include, *e.g.*, Tat (which is a nuclear transcriptional activator protein required for viral replication by HIV type 1), penetratin, Kaposi fibroblast growth factor (FGF) signal peptide sequence, integrin β3 signal peptide sequence, polyarginine peptide Args sequence, Guanine rich-molecular transporters, and sweet arrow peptide. CPPs and methods of using them are described, *e.g.*, in Hällbrink *et al.*, "Prediction of cell-penetrating peptides," *Methods Mol. Biol.*, 2015; 1324:39-58; Ramakrishna *et al.*, "Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA," Genome Res., 2014 June; 24(6):1020-7; and WO 2016205764 A1; each of which is incorporated herein by reference in its entirety.

Various delivery methods for the CRISPR systems described herein are also described, *e.g.*, in U.S. Pat. No. 8,795,965, EP 3009511, WO 2016205764, and WO 2017070605; each of which is incorporated herein by reference in its entirety.

8. Kits

Another aspect of the invention provides a kit, comprising any two or more components of the subject CRISPR/Cas system described herein, such as the Cas13e and Cas13f proteins, derivatives, functional fragments or the various fusions or adducts thereof, guide RNA / crRNA,

complexes thereof, vectors encompassing the same, or host encompassing the same.

In certain embodiments, the kit further comprise an instruction to use the components encompassed therein, and/or instructions for combining with additional components that may be available elsewhere.

In certain embodiments, the kit further comprise one or more nucleotides, such as nucleotide(s) corresponding to those useful to insert the guide RNA coding sequence into a vector and operably linking the coding sequence to one or more control elements of the vector.

In certain embodiments, the kit further comprise one or more buffers that may be used to dissolve any of the components, and/or to provide suitable reaction conditions for one or more of the components. Such buffers may include one or more of PBS, HEPES, Tris, MOPS, Na_2CO_3, $NaHCO_3$, NaB, or combinations thereof. In certain embodiments, the reaction condition includes a proper pH, such as a basic pH. In certain embodiments, the pH is between 7-10.

In certain embodiments, any one or more of the kit components may be stored in a suitable container.

EXAMPLES

Example 1 Identification of Novel Cas13e and Cas13f Systems

A computational pipeline was used to produce an expanded database of class 2 CRISPR-Cas systems from genomic and metagenomic sources. Genome and metagenome sequences were downloaded from NCBI (Benson *et al.*, 2013; Pruitt *et al.*, 2012), NCBI whole genome sequencing (WGS), and DOE JGI Integrated Microbial Genomes (Markowitz *et al.*, 2012). Proteins were predicted (Prodigal (Hyatt *et al.*, 2010) in anon mode) on all contigs at least 5 kb in length, and de-duplicated (*i.e.*, removing identical protein sequences) to construct a complete protein database. Proteins larger than 600 residues were considered as Large Proteins (LPs). Since the currently identified Cas13 proteins are mostly larger than 900 residues in size, in order to reduce the complexity of calculation, only Large Proteins were considered further.

CRISPR arrays were identified using Piler-CR (Edgar, PILER-CR: Fast and accurate identification of CRISPR repeats. *BMC Bioinformatics* 8:18, 2007), using all default parameters. Non-redundant Large Protein sequence-encoding ORFs located within ±10 kb from the CRISPR

arrays were grouped into CRISPR-proximal Large Protein encoding clusters, and the encoded LPs were defined as Cas-LPs.

First, BLASP was used to conduct pairwise alignment between the Cas-LPs, and BLASTP alignment results with Evalue <1E-10 were obtained. MCL was then used to further cluster the Cas-LPs based on the BLASTP results to create families of Cas proteins.

Next, BLASTP was used to align Cas-LPs to all LPs and BLASP alignment results with Evalue<1E-10 were obtained. Cas-LPs families were further expanded according to the BLASTP alignment results. The Cas-LP families were obtained for further analysis with no more than double increase after expansion.

For functional characterization of the candidate Cas proteins, protein family databases Pfam (Finn *et al.*, 2014), NR database, and Cas proteins in NCBI were used to annotate the candidate Cas proteins. Multiple sequence alignment was then conducted for each candidate Cas effector proteins using MAFFT (Katoh and Standley, 2013). JPred and HHpred were then used to analyze conserved regions in these proteins, to identify candidate Cas proteins / families having two conserved RXXXXH motifs.

This analysis led to the identification of seven novel Cas13 effector proteins falling within two new Cas13 families different from all previously identified Class 2 CRISPR-Cas systems. These include Cas13e.1 (SEQ ID NO: 1) and Cas13e.2 (SEQ ID NO: 2) of the new Cas13e family, and Cas13f.1 (SEQ ID NO: 3), Cas13f.2 (SEQ ID NO: 4), Cas13f.3 (SEQ ID NO: 5), Cas13f.4 (SEQ ID NO: 6), and Cas13f.5 (SEQ ID NO: 7) of the new Cas13f family.

MAQVSKQTSKKRELSIDEYQGARKWCFTIAFNKALVNRDKNDGLFVESLLRHEKYSKHDWYDEDTRALIKCSTQAAN
AKAEALRNYFSHYRHSPGCLTFTAEDELRTIMERAYERAIFECRRRETEVIIEFPSLFEGDRITTAGVVFFVSFFVE
RRVLDRLYGAVSGLKKNEGQYKLTRKALSMYCLKDSRFTKAWDKRVLLFRDILAQLGRIPAEAYEYYHGEQGDKKRA
NDNEGTNPKRHKDKFIEFALHYLEAQHSEICFGRRHIVREEAGAGDEHKKHRTKGKVVVDFSKKDEDQSYYISKNNV
IVRIDKNAGPRSYRMGLNELKYLVLLSLQGKGDDAIAKLYRYRQHVENILDVVKVTDKDNHVFLPRFVLEQHGIGRK
AFKQRIDGRVKHVRGVWEKKKAATNEMTLHEKARDILQYVNENCTRSFNPGEYNRLLVCLVGKDVENFQAGLKRLQL
AERIDGRVYSIFAQTSTINEMHQVVCDQILNRLCRIGDQKLYDYVGLGKKDEIDYKQKVAWFKEHISIRRGFLRKKF
WYDSKKGFAKLVEEHLESGGGQRDVGLDKKYYHIDAIGRFEGANPALYETLARDRLCLMMAQYFLGSVRKELGNKIV
WSNDSIELPVEGSVGNEKSIVFSVSDYGKLYVLDDAEFLGRICEYFMPHEKGKIRYHTVYEKGFRAYNDLQKKCVEA
VLAFEEKVVKAKKMSEKEGAHYIDFREILAQTMCKEAEKTAVNKVRRAFFHHHLKFVIDEFGLFSDVMKKYGIEKEW
KFPVK* (SEQ ID NO: 1)

MKVENIKEKSKKAMYLINHYEGPKKWCFAIVLNRACDNYEDNPHLFSKSLLEFEKTSRKDWFDEETRELVEQADTEI
QPNPNLKPNTTANRKLKDIRNYFSHHYHKNECLYFKNDDPIRCIMEAAYEKSKIYIKGKQIEQSDIPLPELFESSGW
ITPAGILLLASFFVERGILHRLMGNIGGFKDNRGEYGLTHDIFTTYCLKGSYSIRAQDHDAVMFRDILGYLSRVPTE
SFQRIKQPQIRKEGQLSERKTDKFITFALNYLEDYGLKDLEGCKACFARSKIVREQENVESINDKEYKPHENKKKVE
IHFDQSKEDRFYINRNNVILKIQKKDGHSNIVRMGVYELKYLVLMSLVGKAKEAVEKIDNYIQDLRDQLPYIEGKNK
EEIKEYVRFFPRFIRSHLGLLQINDEEKIKARLDYVKTKWLDKKEKSKELELHKKGRDILRYINERCDRELNRNVYN
RILELLVSKDLTGFYRELEELKRTRRIDKNIVQNLSGQKTINALHEKVCDLVLKEIESLDTENLRKYLGLIPKEEKE

VTFKEKVDRILKQPVIYKGFLRYQFFKDDKKSFVLLVEDALKEKGGGCDVPLGKEYYKIVSLDKYDKENKTLCETLA
MDRLCLMMARQYYLSLNAKLAQEAQQIEWKKEDSIELIIFTLKNPDQSKQSFSIRFSVRDFTKLYVTDDPEFLARLC
SYFFPVEKEIEYHKLYSEGINKYTNLQKEGIEAILELEKKLIERNRIQSAKNYLSFNEIMNKSGYNKDEQDDLKKVR
NSLLHYKLIFEKEHLKKFYEVMRGEGIEKKWSLIV* (SEQ ID NO: 2)

MNGIELKKEEAAFYFNQAELNLKAIEDNIFDKERRKTLLNNPQILAKMENFIFNFRDVTKNAKGEIDCLLLKLRELR
NFYSHYVHKRDVRELSKGEKPILEKYYQFAIESTGSENVKLEIIENDAWLADAGVLFFLCIFLKKSQANKLISGISG
FKRNDDTGQPRRNLFTYFSIREGYKVVPEMQKHFLLFSLVNHLSNQDDYIEKAHQPYDIGEGLFFHRIASTFLNISG
ILRNMKFYTYQSKRLVEQRGELKREKDIFAWEEPFQGNSYFEINGHKGVIGEDELKELCYAFLIGNQDANKVEGRIT
QFLEKFRNANSVQQVKDDEMLKPEYFPANYFAESGVGRIKDRVLNRLNKAIKSNKAKKGEIIAYDKMREVMAFINNS
LPVDEKLKPKDYKRYLGMVRFWDREKDNIKREFETKEWSKYLPSNFWTAKNLERVYGLAREKNAELFNKLKADVEKM
DERELEKYQKINDAKDLANLRRLASDFGVKWEEKDWDEYSGQIKKQITDSQKLTIMKQRITAGLKKKHGIENLNLRI
TIDINKSRKAVLNRIAIPRGFVKRHILGWQESEKVSKKIREAECEILLSKEYEELSKQFFQSKDYDKMTRINGLYEK
NKLIALMAVYLMGQLRILFKEHTKLDDITKTTVDFKISDKVTVKIPFSNYPSLVYTMSSKYVDNIGNYGFSNKDKDK
PILGKIDVIEKQRMEFIKEVLGFEKYLFDDKIIDKSKFADTATHISFAEIVEELVEKGWDKDRLTKLKDARNKALHG
EILTGTSFDETKSLINELKK* (SEQ ID NO: 3)

MSPDFIKLEKQEAAFYFNQTELNLKAIESNILDKQQRMILLNNPRILAKVGNFIFNFRDVTKNAKGEIDCLLFKLEE
LRNFYSHYVHTDNVKELSNGEKPLLERYYQIAIQATRSEDVKFELFETRNENKITDAGVLFFLCMFLKKSQANKLIS
GISGFKRNDPTGQPRRNLFTYFSAREGYKALPDMQKHFLLFTLVNYLSNQDEYISELKQYGEIGQGAFFNRIASTFL
NISGISGNTKFYSYQSKRIKEQRGELNSEKDSFEWIEPFQGNSYFEINGHKGVIGEDELKELCYALLVAKQDINAVE
GKIMQFLKKFRNTGNLQQVKDDEMLEIEYFPASYFNESKKEDIKKEILGRLDKKIRSCSAKAEKAYDKMKEVMEFIN
NSLPAEEKLKRKDYRRYLKMVRFWSREKGNIEREFRTKEWSKYFSSDFWRKNNLEDVYKLATQKNAELFKNLKAAAE
KMGETEFEKYQQINDVKDLASLRRLTQDFGLKWEEKDWEEYSEQIKKQITDRQKLTIMKQRVTAELKKKHGIENLNL
RITIDSNKSRKAVLNRIAIPRGFVKKHILGWQGSEKISKNIREAECKILLSKKYEELSRQFFEAGNFDKLTQINGLY
EKNKLTAFMSVYLMGRLNIQLNKHTELGNLKKTEVDFKISDKVTEKIPFSQYPSLVYAMSRKYVDNVDKYKFSHQDK
KKPFLGKIDSIEKERIEFIKEVLDFEEYLFKNKVIDKSKFSDTATHISFKEICDEMGKKGCNRNKLTELNNARNAAL
HGEIPSETSFREAKPLINELKK* (SEQ ID NO: 4)

MSPDFIKLEKQEAAFYFNQTELNLKAIESNIFDKQQRVILLNNPQILAKVGDFIFNFRDVTKNAKGEIDCLLLKLRE

LRNFYSHYVYTDDVKILSNGERPLLEKYYQFAIEATGSENVKLEIIESNNRLTEAGVLFFLCMFLKKSQANKLISGI

SGFKRNDPTGQPRRNLFTYFSVREGYKVVPDMQKHFLLFVLVNHLSGQDDYIEKAQKPYDIGEGLFFHRIASTFLNI

SGILRNMEFYIYQSKRLKEQQGELKREKDIFPWIEPFQGNSYFEINGNKGIIGEDELKELCYALLVAGKDVRAVEGK

ITQFLEKFKNADNAQQVEKDEMLDRNNFPANYFAESNIGSIKEKILNRLGKTDDSYNKTGTKIKPYDMMKEVMEFIN

NSLPADEKLKRKDYRRYLKMVRIWDSEKDNIKREFESKEWSKYFSSDFWMAKNLERVYGLAREKNAELFNKLKAVVE

KMDEREFEKYRLINSAEDLASLRRLAKDFGLKWEEKDWQEYSGQIKKQISDRQKLTIMKQRITAELKKKHGIENLNL

RITIDSNKSRKAVLNRIAVPRGFVKEHILGWQGSEKVSKKTREAKCKILLSKEYEELSKQFFQTRNYDKMTQVNGLY

EKNKLLAFMVVYLMERLNILLNKPTELNELEKAEVDFKISDKVMAKIPFSQYPSLVYAMSSKYADSVGSYKFENDEK

NKPFLGKIDTIEKQRMEFIKEVLGFEEYLFEKKIIDKSEFADTATHISFDEICNELIKKGWDKDKLTKLKDARNAAL

HGEIPAETSFREAKPLINGLKK* (SEQ ID NO: 5)

MNIIKLKKEEAAFYFNQTILNLSGLDEIIEKQIPHIISNKENAKKVIDKIFNNRLLLKSVENYIYNFKDVAKNARTE

IEAILLKLVELRNFYSHYVHNDTVKILSNGEKPILEKYYQIAIEATGSKNVKLVIIENNNCLTDSGVLFLLCMFLKK

SQANKLISSVSGFKRNDKEGQPRRNLFTYYSVREGYKVVPDMQKHFLLFALVNHLSEQDDHIEKQQQSDELGKGLFF

HRIASTFLNESGIFNKMQFYTYQSNRLKEKRGELKHEKDTFTWIEPFQGNSYFTLNGHKGVISEDQLKELCYTILIE

KQNVDSLEGKIIQFLKKFQNVSSKQQVDEDELLKREYFPANYFGRAGTGTLKEKILNRLDKRMDPTSKVTDKAYDKM

IEVMEFINMCLPSDEKLRQKDYRRYLKMVRFWNKEKHNIKREFDSKKWTRFLPTELWNKRNLEEAYQLARKENKKKL

EDMRNQVRSLKENDLEKYQQINYVNDLENLRLLSQELGVKWQEKDWVEYSGQIKKQISDNQKLTIMKQRITAELKKM

HGIENLNLRISIDTNKSRQTVMNRIALPKGFVKNHIQQNSSEKISKRIREDYCKIELSGKYEELSRQFFDKKNFDKM

TLINGLCEKNKLIAFMVIYLLERLGFELKEKTKLGELKQTRMTYKISDKVKEDIPLSYYPKLVYAMNRKYVDNIDSY
AFAAYESKKAILDKVDIIEKQRMEFIKQVLCFEEYIFENRIIEKSKFNDEETHISFTQIHDELIKKGRDTEKLSKLK
HARNKALHGEIPDGTSFEKAKLLINEIKK* (SEQ ID NO: 6)

MNAIELKKEEAAFYFNQARLNISGLDEIIEKQLPHIGSNRENAKKTVDMILDNPEVLKKMENYVFNSRDIAKNARGE
LEALLLKLVELRNFYSHYVHKDDVKTLSYGEKPLLDKYYEIAIEATGSKDVRLEIIDDKNKLTDAGVLFLLCMFLKK
SEANKLISSIRGFKRNDKEGQPRRNLFTYYSVREGYKVVPDMQKHFLLFTLVNHLSNQDEYISNLRPNQEIGQGGFF
HRIASKFLSDSGILHSMKFYTYRSKRLTEQRGELKPKKDHFTWIEPFQGNSYFSVQGQKGVIGEEQLKELCYVLLVA
REDFRAVEGKVTQFLKKFQNANNVQQVEKDEVLEKEYFPANYFENRDVGRVKDKILNRLKKITESYKAKGREVKAYD
KMKEVMEFINNCLPTDENLKLKDYRRYLKMVRFWGREKENIKREFDSKKWERFLPRELWQKRNLEDAYQLAKEKNTE
LFNKLKTTVERMNELEFEKYQQINDAKDLANLRQLARDFGVKWEEKDWQEYSGQIKKQITDRQKLTIMKQRITAALK
KKQGIENLNLRITTDTNKSRKVVLNRIALPKGFVRKHILKTDIKISKQIRQSQCPIILSNNYMKLAKEFFEERNFDK
MTQINGLFEKNVLIAFMIVYLMEQLNLRLGKNTELSNLKKTEVNFTITDKVTEKVQISQYPSLVFAINREYVDGISG
YKLPPKKPKEPPYTFFEKIDAIEKERMEFIKQVLGFEEHLFEKNVIDKTRFTDTATHISFNEICDELIKKGWDENKI
IKLKDARNAALHGKIPEDTSFDEAKVLINELKK* (SEQ ID NO: 7)

DNA encoding the corresponding Direct Repeat (DR) sequences in the respective pre-crRNA sequences are SEQ ID NOs: 8-14, respectively.

GCTGGAGCAGCCCCCGATTTGTGGGGTGATTACAGC (SEQ ID NO: 8)

GCTGAAGAAGCCTCCGATTTGAGAGGTGATTACAGC (SEQ ID NO: 9)

GCTGTGATAGACCTCGATTTGTGGGGTAGTAACAGC (SEQ ID NO: 10)

GCTGTGATAGACCTCGATTTGTGGGGTAGTAACAGC (SEQ ID NO: 11)

GCTGTGATAGACCTCGATTTGTGGGGTAGTAACAGC (SEQ ID NO: 12)

GCTGTGATGGGCCTCAATTTGTGGGGAAGTAACAGC (SEQ ID NO: 13)

GCTGTGATAGGCCTCGATTTGTGGGGTAGTAACAGC (SEQ ID NO: 14)

Natural (wild-type) DNA coding sequences for Cas13e.1, Cas13e.2, Cas13f.1, Cas13f.2, Cas13f.3, Cas13f.4, and Cas13f.5 proteins are SEQ ID NOs: 15-21, respectively.

ATGGCGCAAGTGTCAAAGCAGACTTCGAAAAAGAGAGAGTTGTCTATCGATGAATATCAAGGTGCTCGGAAATGGTG
TTTTACGATTGCCTTCAACAAGGCTCTTGTGAATCGAGATAAGAACGACGGGCTTTTTGTCGAGTCGCTGTTACGCC
ATGAAAAGTATTCAAAGCACGACTGGTACGATGAGGATACACGCGCTTTGATCAAGTGTAGCACACAAGCGGCCAAT
GCGAAGGCCGAGGCGTTAAGAAACTATTTCTCCCACTATCGACATTCGCCCGGGTGTCTGACATTTACAGCAGAAGA
TGAGTTGCGGACAATCATGGAAAGGGCGTATGAGCGGGCGATCTTTGAATGCAGGAGACGCGAAACTGAAGTGATCA
TCGAGTTTCCCAGCCTGTTCGAAGGCGACCGGATCACTACGGCGGGGGTTGTGTTTTTCGTTTCGTTCTTTGTTGAA
CGGCGGGTGCTGGATCGTTTGTACGGTGCGGTAAGTGGGCTTAAGAAAAACGAAGGACAGTACAAGCTGACTCGGAA
GGCGCTTTCGATGTATTGCCTGAAAGACAGTCGTTTCACGAAGGCGTGGGACAAACGCGTGCTGCTTTTCAGGGATA
TACTCGCGCAGCTTGGACGCATCCCTGCGGAGGCGTATGAATACTACCACGGAGAGCAGGGCGACAAGAAAAGAGCA
AACGACAATGAGGGGACGAATCCGAAACGCCATAAAGACAAGTTCATCGAGTTTGCACTGCATTATCTGGAGGCGCA
ACACAGTGAGATATGCTTCGGGCGGCGACACATTGTCAGGGAGGAGGCCGGGGCAGGCGACGAACACAAAAAGCACA
GGACCAAAGGCAAGGTAGTTGTCGACTTTTCAAAAAAAGACGAAGATCAGTCATACTATATCAGTAAGAACAATGTT
ATCGTCAGGATTGATAAGAATGCCGGGCCTCGGAGTTATCGCATGGGGCTTAACGAATTGAAATACCTTGTATTGCT
TAGCCTTCAGGGAAAGGGCGACGATGCGATTGCAAAACTGTACAGGTATCGGCAGCATGTGGAGAACATTCTGGATG
TAGTGAAGGTCACAGATAAGGATAATCACGTCTTCCTGCCGCGATTTGTGCTGGAGCAACATGGGATTGGCAGGAAA

GCTTTTAAGCAAAGAATAGACGGCAGAGTAAAGCATGTTCGAGGGGTGTGGGAAAAGAAGAAGGCGGCGACCAACGA
GATGACACTTCACGAGAAGGCGCGGGACATTCTTCAATACGTAAATGAAAATTGCACGAGGTCTTTCAATCCCGGCG
AGTACAACCGGCTGCTGGTGTGTCTGGTTGGCAAGGATGTTGAGAATTTTCAGGCGGGACTGAAACGCCTGCAACTG
GCCGAGCGAATCGACGGGCGGGTATATTCAATTTTTGCGCAGACCTCCACAATAAACGAGATGCATCAGGTGGTGTG
TGATCAGATTCTCAACAGACTTTGCCGAATCGGCGATCAGAAGCTCTACGATTATGTGGGGCTTGGGAAGAAGGATG
AAATAGATTACAAGCAGAAGGTTGCATGGTTCAAGGAGCATATTTCTATCCGCAGGGGTTTCTTGCGCAAGAAGTTC
TGGTATGACAGCAAGAAGGGATTCGCGAAGCTTGTGGAAGAGCATTTGGAAAGCGGCGGCGGACAGAGGGACGTTGG
GCTGGATAAAAAGTATTATCATATTGATGCGATTGGGCGATTCGAGGGTGCTAATCCAGCCTTGTATGAAACGCTGG
CGCGAGACCGTTTGTGTCTGATGATGGCGCAATACTTCCTGGGGAGTGTACGCAAGGAATTGGGTAATAAAATTGTG
TGGTCGAATGATAGCATCGAGTTGCCCGTGGAGGGCTCAGTGGGTAACGAAAAAAGCATCGTCTTCTCAGTGAGTGA
TTACGGCAAGTTATATGTGTTGGATGACGCTGAGTTTCTTGGGCGGATATGTGAGTACTTTATGCCGCACGAAAAAG
GGAAGATACGGTATCATACAGTTTACGAAAAAGGGTTTAGGGCATATAATGATCTGCAGAAGAAATGTGTCGAGGCG
GTGCTGGCGTTTGAAGAGAAGGTTGTCAAAGCCAAAAAGATGAGCGAGAAGGAAGGGGCGCATTATATTGATTTTCG
TGAGATACTGGCACAAACAATGTGTAAAGAGGCGGAGAAGACCGCCGTGAATAAGGTGCGTAGAGCGTTTTTCCATC
ATCATTTAAAGTTTGTGATAGATGAATTTGGGTTGTTTAGTGATGTTATGAAGAAATATGGAATTGAAAAGGAGTGG
AAGTTTCCTGTTAAATGA (SEQ ID NO: 15)

ATGAAGGTTGAAAATATTAAAGAAAAAAGCAAAAAAGCAATGTATTTAATCAACCATTATGAGGGACCCAAAAAATG
GTGTTTTGCAATAGTTCTGAATAGGGCATGTGATAATTACGAGGACAATCCACACTTGTTTTCCAAATCACTTTTGG
AATTTGAAAAAACAAGTCGAAAAGATTGGTTTGACGAAGAAACACGAGAGCTTGTTGAGCAAGCAGATACAGAAATA
CAGCCAAATCCTAACCTGAAACCTAATACAACAGCTAACCGAAAACTCAAAGATATAAGAAACTATTTTTCGCATCA
TTATCACAAGAACGAATGCCTGTATTTTAAGAACGATGATCCCATACGCTGCATTATGGAAGCGGCGTATGAAAAAT
CTAAAATTTATATCAAAGGAAAGCAGATTGAGCAAAGCGATATACCATTGCCCGAATTGTTTGAAAGCAGCGGTTGG
ATTACACCGGCGGGGATTTTGTTACTGGCATCCTTTTTTGTTGAACGAGGGATTCTACATCGCTTGATGGGAAATAT
CGGAGGATTTAAAGATAATCGAGGCGAATACGGTCTTACACACGATATTTTTACCACCTATTGTCTTAAGGGTAGTT
ATTCAATTCGGGCGCAGGATCATGATGCGGTAATGTTCAGAGATATTCTCGGCTATCTGTCACGAGTTCCCACTGAG
TCATTTCAGCGTATCAAGCAACCTCAAATACGAAAAGAAGGCCAATTAAGTGAAAGAAAGACGGACAAATTTATAAC
ATTTGCACTAAATTATCTTGAGGATTATGGGCTGAAAGATTTGGAAGGCTGCAAAGCCTGTTTTGCCAGAAGTAAAA
TTGTAAGGGAACAAGAAAATGTTGAAAGCATAAATGATAAGGAATACAAACCTCACGAGAACAAAAAGAAAGTTGAA
ATTCACTTCGATCAGAGCAAAGAAGACCGATTTTATATTAATCGCAATAACGTTATTTTGAAGATTCAGAAGAAAGA
TGGACATTCCAACATAGTTAGGATGGGAGTATATGAACTTAAATATCTCGTTCTTATGAGTTTAGTGGGAAAAGCAA
AAGAAGCAGTTGAAAAAATTGACAACTATATCCAGGATTTGCGAGACCAGTTGCCTTACATAGAGGGGAAAAATAAG
GAAGAGATTAAAGAATACGTCAGGTTCTTTCCACGATTTATACGTTCTCACCTCGGTTTACTACAGATTAACGATGA
AGAAAAGATAAAAGCTCGATTAGATTATGTTAAGACCAAGTGGTTAGATAAAAAGGAAAAATCGAAAGAGCTTGAAC
TTCATAAAAAAGGACGGGACATCCTCAGGTATATCAACGAGCGATGTGATAGAGAGCTTAACAGGAATGTATATAAC
CGTATTTTAGAGCTCCTGGTCAGCAAAGACCTCACTGGTTTTTATCGTGAGCTTGAAGAACTAAAAAGAACAAGGCG
GATAGATAAAAATATTGTCCAGAATCTTTCTGGGCAAAAAACCATTAATGCACTGCATGAAAAGGTCTGTGATCTGG
TGCTGAAGGAAATCGAAAGTCTCGATACAGAAAATCTCAGGAAATATCTTGGATTGATACCCAAAGAAGAAAAAGAG
GTCACTTTCAAAGAAAAGGTCGATAGGATTTTGAAACAGCCAGTTATTTACAAAGGGTTTCTGAGATACCAATTCTT
CAAAGATGACAAAAAGAGTTTTGTCTTACTTGTTGAAGACGCATTGAAGGAAAAAGGAGGAGGTTGTGATGTTCCTC
TTGGGAAAGAGTATTATAAAATCGTGTCACTTGATAAGTATGATAAAGAAAATAAAACCCTGTGTGAAACTCTGGCG
ATGGATAGGCTTTGCCTTATGATGGCAAGACAATATTATCTCAGTCTGAATGCAAAACTTGCACAGGAAGCTCAGCA
AATCGAATGGAAGAAAGAAGATAGTATAGAATTGATTATTTTCACCTTAAAAAATCCCGATCAATCAAAGCAGAGTT
TTTCTATACGGTTTTCGGTCAGAGATTTTACGAAGTTGTATGTAACGGATGATCCTGAATTTCTGGCCCGGCTTTGT
TCCTACTTTTTCCCAGTTGAAAAAGAGATTGAATATCACAAGCTCTATTCAGAAGGGATAAATAAATACACAAACCT
GCAAAAAGAGGGAATCGAAGCAATACTCGAGCTTGAAAAAAAGCTTATTGAACGAAATCGGATTCAATCTGCAAAAA
ATTATCTCTCATTTAATGAGATAATGAATAAAAGCGGTTATAATAAAGATGAGCAGGATGATCTAAAGAAGGTGCGA
AATTCTCTTTTGCATTATAAGCTTATCTTTGAGAAAGAACATCTCAAGAAGTTCTATGAGGTTATGAGAGGAGAAGG
GATAGAGAAAAAGTGGTCTTTAATAGTATGA (SEQ ID NO: 16)

ATGAATGGCATTGAATTAAAAAAAGAAGAAGCAGCATTTTATTTTAATCAGGCAGAGCTTAATTTAAAAGCCATAGA
AGACAATATTTTTGATAAAGAAAGACGAAAGACTCTGCTTAATAATCCACAGATACTTGCCAAAATGGAAAATTTCA
TTTTCAATTTCAGAGATGTAACAAAAAATGCAAAAGGGGAAATTGACTGCTTGCTGTTGAAACTAAGAGAGCTGAGA
AACTTTTACTCGCATTATGTCCACAAACGAGATGTAAGAGAATTAAGCAAGGGCGAGAAACCTATACTTGAAAAGTA
TTACCAATTTGCGATTGAATCAACCGGAAGTGAAAATGTTAAACTTGAGATAATAGAAAACGACGCGTGGCTTGCAG
ATGCCGGTGTGTTGTTTTTCTTATGTATTTTTTTGAAGAAATCTCAGGCAAATAAGCTTATAAGCGGTATCAGCGGT
TTTAAAAGAAACGATGATACCGGTCAGCCGAGAAGGAATTTATTTACCTATTTCAGTATAAGGGAGGGATACAAGGT
TGTTCCGGAAATGCAGAAACATTTCCTTTTGTTTTCTCTTGTTAATCATCTCTCTAATCAAGATGATTATATTGAAA

AAGCGCATCAGCCATACGATATAGGCGAGGGTTTATTTTTTCATCGAATAGCTTCTACATTTCTTAATATAAGTGGG
ATTTTAAGAAATATGAAATTCTATACCTATCAGAGTAAAAGGTTAGTAGAGCAGCGGGGAGAACTCAAACGAGAAAA
GGATATTTTTGCGTGGGAAGAACCGTTTCAAGGAAATAGTTATTTTGAAATAAATGGTCATAAAGGAGTAATCGGTG
AAGATGAATTGAAGGAACTATGTTATGCATTTCTGATTGGCAATCAAGATGCTAATAAAGTGGAAGGCAGGATTACA
CAATTTCTAGAAAAGTTTAGAAATGCGAACAGTGTGCAACAAGTTAAAGATGATGAAATGCTAAAACCAGAGTATTT
TCCTGCAAATTATTTTGCTGAATCAGGCGTCGGAAGAATAAAGGATAGAGTGCTTAATCGTTTGAATAAAGCGATTA
AAAGCAATAAGGCCAAGAAAGGAGAGATTATAGCATACGATAAGATGAGAGAGGTTATGGCGTTCATAAATAATTCT
CTGCCGGTAGATGAAAAATTGAAACCAAAAGATTACAAACGATATCTGGGAATGGTTCGTTTCTGGGACAGGGAAAA
AGATAACATAAAGCGGGAGTTCGAGACAAAAGAATGGTCTAAATATCTTCCATCTAATTTCTGGACGGCAAAAAACC
TTGAAAGGGTCTATGGTCTGGCAAGAGAGAAAAACGCAGAATTATTCAATAAACTAAAAGCGGATGTAGAAAAAATG
GACGAACGGGAACTTGAGAAGTATCAGAAGATAAATGATGCAAAGGATTTGGCAAATTTACGCCGGCTTGCAAGCGA
CTTTGGTGTGAAGTGGGAAGAAAAAGACTGGGATGAGTATTCAGGACAGATAAAAAAACAAATTACAGACAGCCAGA
AACTAACAATAATGAAGCAGCGGATAACCGCAGGACTAAAGAAAAAGCACGGCATAGAAAATCTTAACCTGAGAATA
ACTATCGACATCAATAAAAGCAGAAAGGCAGTTTTGAACAGAATTGCGATTCCGAGGGGTTTTGTAAAAAGGCATAT
TTTAGGATGGCAAGAGTCTGAGAAGGTATCGAAAAAGATAAGAGAGGCAGAATGCGAAATTCTGCTGTCGAAAGAAT
ACGAAGAACTATCGAAACAATTTTTCCAAAGCAAAGATTATGACAAAATGACACGGATAAATGGCCTTTATGAAAAA
AACAAACTTATAGCCCTGATGGCAGTTTATCTAATGGGGCAATTGAGAATCCTGTTTAAAGAACACACAAAACTTGA
CGATATTACGAAAACAACTGTGGATTTCAAAATATCTGATAAGGTGACGGTAAAAATCCCCTTTTCAAATTATCCTT
CGCTCGTTTATACAATGTCCAGTAAGTATGTTGATAATATAGGGAATTATGGATTTTCCAACAAAGATAAAGACAAG
CCGATTTTAGGTAAGATTGATGTAATAGAAAAACAGCGAATGGAATTTATAAAAGAGGTTCTTGGTTTTGAAAAATA
TCTTTTTGATGATAAAATAATAGATAAAAGCAAATTTGCTGATACAGCGACTCATATAAGTTTTGCAGAAATAGTTG
AGGAGCTTGTTGAAAAAGGATGGGACAAAGACAGACTGACAAAACTTAAAGATGCAAGAAATAAAGCCCTGCATGGT
GAAATACTGACGGGAACCAGCTTTGATGAAACAAAATCATTGATAAACGAATTAAAAAAATGA (SEQ ID NO:
17)

ATGTCCCCAGATTTCATCAAATTAGAAAAACAGGAAGCAGCTTTTTACTTTAATCAGACAGAGCTTAATTTAAAAGC
CATAGAAAGCAATATTTTAGACAAACAACAGCGAATGATTCTGCTTAATAATCCACGGATACTTGCCAAAGTAGGAA
ATTTCATTTTCAATTTCAGAGATGTAACAAAAAATGCAAAAGGAGAAATAGACTGTCTGCTATTTAAACTGGAAGAG
CTAAGAAACTTTTACTCGCATTATGTTCATACCGACAATGTAAAGGAATTGAGTAACGGAGAAAAACCCCTACTGGA
AAGATATTATCAAATCGCTATTCAGGCAACCAGGAGTGAGGATGTTAAGTTCGAATTGTTTGAAACAAGAAACGAGA
ATAAGATTACGGATGCCGGTGTATTGTTTTTCTTATGTATGTTTTTAAAAAAATCACAGGCAAACAAGCTTATAAGC
GGTATCAGCGGCTTCAAAAGAAATGATCCAACAGGCCAGCCGAGAAGAAACTTATTTACCTATTTCAGTGCAAGAGA
AGGATATAAGGCTTTGCCTGATATGCAGAAACATTTTCTTCTTTTTACTCTGGTTAATTATTTGTCGAATCAGGATG
AGTATATCAGCGAGCTTAAACAATATGGAGAGATTGGTCAAGGAGCCTTTTTTAATCGAATAGCTTCAACATTTTTG
AATATCAGCGGGATTTCAGGAAATACGAAATTCTATTCGTATCAAAGTAAAAGGATAAAAGAGCAGCGAGGCGAACT
CAATAGCGAAAAGGACAGCTTTGAATGGATAGAGCCTTTCCAAGGAAACAGCTATTTTGAAATAAATGGGCATAAAG
GAGTAATCGGCGAAGACGAATTAAAAGAACTTTGTTATGCATTGTTGGTTGCCAAGCAAGATATTAATGCCGTTGAA
GGCAAAATTATGCAATTCCTGAAAAAGTTTAGAAATACTGGCAATTTGCAGCAAGTTAAAGATGATGAAATGCTGGA
AATAGAATATTTTCCCGCAAGTTATTTTAATGAATCAAAAAAAGAGGACATAAAGAAAGAGATTCTTGGCCGGCTGG
ATAAAAAGATTCGCTCCTGCTCTGCAAAGGCAGAAAAAGCCTATGATAAGATGAAAGAGGTGATGGAGTTTATAAAT
AATTCTCTGCCGGCAGAGGAAAAATTGAAACGCAAAGATTATAGAAGATATCTAAAGATGGTTCGTTTCTGGAGCAG
AGAAAAAGGCAATATAGAGCGGGAATTTAGAACAAAGGAATGGTCAAAATATTTTTCATCTGATTTTTGGCGGAAGA
ACAATCTTGAAGATGTGTACAAACTGGCAACACAAAAAAACGCTGAACTGTTCAAAAATCTAAAAGCGGCAGCAGAG
AAAATGGGTGAAACGGAATTTGAAAAGTATCAGCAGATAAACGATGTAAAGGATTTGGCAAGTTTAAGGCGGCTTAC
GCAAGATTTTGGTTTGAAGTGGGAAGAAAAGGACTGGGAGGAGTATTCCGAGCAGATAAAAAAACAAATTACGGACA
GGCAGAAACTGACAATAATGAAACAAAGGGTTACGGCTGAACTAAAGAAAAAGCACGGCATAGAAAATCTTAATCTG
AGAATAACCATCGACAGCAATAAAAGCAGAAAGGCGGTTTTGAACAGAATAGCAATTCCAAGAGGATTTGTAAAAAA
ACATATTTTAGGCTGGCAGGGATCTGAGAAGATATCGAAAAATATAAGGGAAGCAGAATGCAAAATTCTGCTATCGA
AAAAATATGAAGAGTTATCAAGGCAGTTTTTTGAAGCCGGTAATTTCGATAAGCTGACGCAGATAAATGGTCTTTAT
GAAAAGAATAAACTTACAGCTTTTATGTCAGTATATTTGATGGGTCGGTTGAATATTCAGCTTAATAAGCACACAGA
ACTTGGAAATCTTAAAAAAACAGAGGTGGATTTTAAGATATCTGATAAGGTGACTGAAAAAATACCGTTTTCTCAGT
ATCCTTCGCTTGTCTATGCGATGTCTCGCAAATATGTTGACAATGTGGATAAATATAAATTTTCTCATCAAGATAAA
AAGAAGCCATTTTTAGGTAAAATTGATTCAATTGAAAAAGAACGTATTGAATTCATAAAAGAGGTTCTCGATTTTGA
AGAGTATCTTTTTAAAAATAAGGTAATAGATAAAAGCAAATTTTCCGATACAGCGACTCATATTAGCTTTAAGGAAA
TATGTGATGAAATGGGTAAAAAAGGATGTAACCGAAACAAACTAACCGAACTTAACAACGCAAGGAACGCAGCCCTG
CATGGTGAAATACCGTCGGAGACCTCTTTTCGTGAAGCAAAACCGTTGATAAATGAATTGAAAAAATGA (SEQ ID
NO: 18)

ATGTCCCCAGATTTCATCAAATTAGAAAAACAAGAAGCAGCTTTTTACTTTAATCAGACAGAGCTTAATTTAAAAGC
CATAGAAAGCAATATTTTCGACAAACAACAGCGAGTGATTCTGCTTAATAATCCACAGATACTTGCCAAAGTAGGAG
ATTTTATTTTCAATTTCAGAGATGTAACAAAAAACGCAAAAGGAGAAATAGACTGTTTGCTATTGAAACTAAGAGAG
CTGAGAAACTTTTACTCACACTATGTCTATACCGATGACGTGAAGATATTGAGTAACGGCGAAAGACCTCTGCTGGA
AAAATATTATCAATTTGCGATTGAAGCAACCGGAAGTGAAAATGTTAAACTTGAAATAATAGAAAGCAACAACCGAC
TTACGGAAGCGGGCGTGCTGTTTTTCTTGTGTATGTTTTTGAAAAAGTCTCAGGCAAATAAGCTTATAAGCGGTATC
AGCGGTTTTAAAAGAAATGACCCGACAGGTCAGCCGAGAAGGAATTTATTTACCTACTTCAGTGTAAGGGAGGGATA
CAAGGTTGTGCCGGATATGCAGAAACATTTTCTTTTGTTTGTTCTTGTCAATCATCTCTCTGGTCAGGATGATTATA
TTGAAAAGGCGCAAAAGCCATACGATATAGGCGAGGGTTTATTTTTTCATCGAATAGCTTCTACATTTCTTAATATC
AGTGGGATTTTAAGAAATATGGAATTCTATATTTACCAGAGCAAAAGACTAAAGGAGCAGCAAGGAGAGCTCAAACG
TGAAAAGGATATTTTTCCATGGATAGAGCCTTTCCAGGGAAATAGTTATTTTGAAATAAATGGTAATAAAGGAATAA
TCGGCGAAGATGAATTGAAAGAGCTTTGTTATGCGTTGCTGGTTGCAGGAAAAGATGTCAGAGCCGTCGAAGGTAAA
ATAACACAATTTTTGGAAAAGTTTAAAAATGCGGACAATGCTCAGCAAGTTGAAAAAGATGAAATGCTGGACAGAAA
CAATTTTCCCGCCAATTATTTCGCCGAATCGAACATCGGCAGCATAAAGGAAAAAATACTTAATCGTTTGGGAAAAA
CTGATGATAGTTATAATAAGACGGGGACAAAGATTAAACCATACGACATGATGAAAGAGGTAATGGAGTTTATAAAT
AATTCTCTTCCGGCAGATGAAAAATTGAAACGCAAAGATTACAGAAGATATCTAAAGATGGTTCGTATCTGGGACAG
TGAGAAAGATAATATAAAGCGGGAGTTTGAAAGCAAAGAATGGTCAAAATATTTTTCATCTGATTTCTGGATGGCAA
AAAATCTTGAAAGGGTCTATGGGTTGGCAAGAGAGAAAAACGCCGAATTATTCAATAAGCTAAAAGCGGTTGTGGAG
AAAATGGACGAGCGGGAATTTGAGAAGTATCGGCTGATAAATAGCGCAGAGGATTTGGCAAGTTTAAGACGGCTTGC
GAAAGATTTTGGCCTGAAGTGGGAAGAAAAGGACTGGCAAGAGTATTCTGGGCAGATAAAAAAACAAATTTCTGACA
GGCAGAAACTGACAATAATGAAACAAAGGATTACGGCTGAACTAAAGAAAAAGCACGGCATAGAAAATCTCAATCTT
AGAATAACCATCGACAGCAATAAAAGCAGAAAGGCAGTTTTGAACAGAATCGCAGTTCCAAGAGGTTTTGTGAAAGA
GCATATTTTAGGATGGCAGGGGTCTGAGAAGGTATCGAAAAAGACAAGAGAAGCAAAGTGCAAAATTCTGCTCTCGA
AAGAATATGAAGAATTATCAAAGCAATTTTTCCAAACCAGAAATTACGACAAGATGACGCAGGTAAACGGTCTTTAC
GAAAAGAATAAACTCTTAGCATTTATGGTCGTTTATCTTATGGAGCGGTTGAATATCCTGCTTAATAAGCCCACAGA
ACTTAATGAACTTGAAAAAGCAGAGGTGGATTTCAAGATATCTGATAAGGTGATGGCCAAAATCCCGTTTTCACAGT
ATCCTTCGCTTGTGTACGCGATGTCCAGCAAATATGCTGATAGTGTAGGCAGTTATAAATTTGAGAATGATGAAAAA
AACAAGCCGTTTTTAGGCAAGATCGATACAATAGAAAAACAACGAATGGAGTTTATAAAAGAAGTCCTTGGTTTTGA
AGAGTATCTTTTTGAAAAGAAGATAATAGATAAAAGCGAATTTGCCGACACAGCGACTCATATAAGTTTTGATGAAA
TATGTAATGAGCTTATTAAAAAAGGATGGGATAAAGACAAACTAACCAAACTTAAAGATGCCAGGAACGCGGCCCTG
CATGGCGAAATACCGGCGGAGACCTCTTTTCGTGAAGCAAAACCGTTGATAAATGGATTGAAAAAATGA (SEQ ID
NO: 19)

ATGAACATCATTAAATTAAAAAAAGAAGAAGCTGCGTTTTATTTTAATCAGACGATCCTCAATCTTTCAGGGCTTGA
TGAAATTATTGAAAAACAAATTCCGCACATAATCAGCAACAAGGAAAATGCAAAGAAAGTGATTGATAAGATTTTCA
ATAACCGCTTATTATTAAAAAGTGTGGAGAATTATATCTACAACTTTAAAGATGTGGCTAAAAACGCAAGAACTGAA
ATTGAGGCTATATTGTTGAAATTAGTAGAGCTACGTAATTTTTACTCACATTACGTTCATAATGATACCGTCAAGAT
ACTAAGTAACGGTGAAAAACCTATACTGGAAAAATATTATCAAATTGCTATAGAAGCAACCGGAAGTAAAAATGTTA
AACTTGTAATCATAGAAAACAACAACTGTCTCACGGATTCTGGCGTGCTGTTTTTGCTGTGTATGTTCTTAAAAAAA
TCACAGGCAAACAAGCTTATAAGTTCCGTTAGTGGTTTTAAAAGGAATGATAAAGAAGGACAACCGAGAAGAAATCT
ATTCACTTATTATAGTGTGAGGGAGGGATATAAGGTTGTGCCTGATATGCAGAAGCATTTCCTTCTATTCGCTCTGG
TCAATCATCTATCTGAGCAGGATGATCATATTGAGAAGCAGCAGCAGTCAGACGAGCTCGGTAAGGGTTTGTTTTTC
CATCGTATAGCTTCGACTTTTTTAAACGAGAGCGGCATCTTCAATAAAATGCAATTTTATACATATCAGAGCAACAG
GCTAAAAGAGAAAAGAGGAGAACTCAAACACGAAAAGGATACCTTTACATGGATAGAGCCTTTTCAAGGCAATAGTT
ATTTTACGTTAAATGGACATAAGGGAGTGATTAGTGAAGATCAATTGAAGGAGCTTTGTTACACAATTTTAATTGAG
AAGCAAAACGTTGATTCCTTGGAAGGTAAAATTATACAATTTCTCAAAAAATTTCAGAATGTCAGCAGCAAGCAGCA
AGTTGACGAAGATGAATTGCTTAAAAGAGAATATTTCCCTGCAAATTACTTTGGCCGGGCAGGAACAGGGACCCTAA
AAGAAAAGATTCTAAACCGGCTTGATAAGAGGATGGATCCTACATCTAAAGTGACGGATAAAGCTTATGACAAAATG
ATTGAAGTGATGGAATTTATCAATATGTGCCTTCCGTCTGATGAGAAGTTGAGGCAAAAGGATTATAGACGATACTT
AAAGATGGTTCGTTTCTGGAATAAGGAAAAGCATAACATTAAGCGCGAGTTTGACAGTAAAAAATGGACGAGGTTTT
TGCCGACGGAATTGTGGAATAAAAGAAATCTAGAAGAAGCCTATCAATTAGCACGGAAAGAGAACAAAAAGAAACTT
GAAGATATGAGAAATCAAGTACGAAGCCTTAAAGAAAATGACCTTGAAAAATATCAGCAGATTAATTACGTTAATGA
CCTGGAGAATTTAAGGCTTCTGTCACAGGAGTTAGGTGTGAAATGGCAGGAAAAGGACTGGGTTGAATATTCCGGGC
AGATAAAGAAGCAGATATCAGACAATCAGAAACTTACAATCATGAAACAAAGGATTACCGCTGAACTAAAGAAAATG
CACGGCATCGAGAATCTTAATCTTAGAATAAGCATTGACACGAATAAAAGCAGGCAGACGGTTATGAACAGGATAGC
TTTGCCCAAAGGTTTTGTGAAGAATCATATCCAGCAAAATTCGTCTGAGAAAATATCGAAAAGAATAAGAGAGGATT
ATTGTAAAATTGAGCTATCGGGAAAATATGAAGAACTTTCAAGGCAATTTTTTGATAAAAAGAATTTCGATAAGATG

ACACTGATAAACGGCCTTTGTGAAAAGAACAAACTTATCGCATTTATGGTTATCTATCTTTTGGAGCGGCTTGGATT
TGAATTAAAGGAGAAAACAAAATTAGGCGAGCTTAAACAAACAAGGATGACATATAAAATATCCGATAAGGTAAAAG
AAGATATCCCGCTTTCCTATTACCCCAAGCTTGTGTATGCAATGAACCGAAAATATGTTGACAATATCGATAGTTAT
GCATTTGCGGCTTACGAATCCAAAAAAGCTATTTTGGATAAAGTGGATATCATAGAAAAGCAACGTATGGAATTTAT
CAAACAAGTTCTCTGTTTTGAGGAATATATTTTCGAAAATAGGATTATCGAAAAAAGCAAATTTAATGACGAGGAGA
CTCATATAAGTTTTACACAAATACATGATGAGCTTATTAAAAAAGGACGGGACACAGAAAAACTCTCTAAACTCAAA
CATGCAAGGAATAAAGCCTTGCACGGCGAGATTCCTGATGGGACTTCTTTTGAAAAAGCAAAGCTATTGATAAATGA
AATCAAAAAATGA (SEQ ID NO: 20)

ATGAATGCTATCGAACTAAAAAAAGAGGAAGCAGCATTTTATTTTAATCAGGCAAGACTCAACATTTCAGGACTTGA
TGAAATTATTGAAAAGCAGTTACCACATATAGGTAGTAACAGGGAGAATGCGAAAAAAACTGTTGATATGATTTTGG
ATAATCCCGAAGTCTTGAAGAAGATGGAAAATTATGTCTTTAACTCACGAGATATAGCAAAGAACGCAAGAGGTGAA
CTTGAAGCATTGTTGTTGAAATTAGTAGAACTGCGTAATTTTTATTCACATTATGTTCATAAAGATGATGTTAAGAC
ATTGAGTTACGGAGAAAAACCTTTACTGGATAAATATTATGAAATTGCGATTGAAGCGACCGGAAGTAAAGATGTCA
GACTTGAGATAATAGATGATAAAAATAAGCTTACAGATGCCGGTGTGCTTTTTTTATTGTGTATGTTTTTGAAAAAA
TCAGAGGCAAACAAACTTATCAGTTCAATCAGGGGCTTTAAAAGAAACGATAAAGAAGGCCAGCCGAGAAGAAATCT
ATTCACTTACTACAGTGTCAGAGAGGGATATAAGGTTGTGCCTGATATGCAGAAACATTTTCTTTTATTCACACTGG
TTAACCATTTGTCAAATCAGGATGAATACATCAGTAATCTTAGGCCGAATCAAGAAATCGGCCAAGGGGGATTTTTC
CATAGAATAGCATCAAAATTTTTGAGCGATAGCGGGATTTTACATAGTATGAAATTCTACACCTACCGGAGTAAAAG
ACTAACAGAACAACGGGGGGAGCTTAAGCCGAAAAAAGATCATTTTACATGGATAGAGCCTTTTCAGGGAAACAGTT
ATTTTTCAGTGCAGGGCCAAAAAGGAGTAATTGGTGAAGAGCAATTAAAGGAGCTTTGTTATGTATTGCTGGTTGCC
AGAGAAGATTTTAGGGCCGTTGAGGGCAAAGTTACACAATTTCTGAAAAAGTTTCAGAATGCTAATAACGTACAGCA
AGTTGAAAAAGATGAAGTGCTGGAAAAAGAATATTTTCCTGCAAATTATTTTGAAAATCGAGACGTAGGCAGAGTAA
AGGATAAGATACTTAATCGTTTGAAAAAAATCACTGAAAGCTATAAAGCTAAAGGGAGGGAGGTTAAAGCCTATGAC
AAGATGAAAGAGGTAATGGAGTTTATAAATAATTGCCTGCCAACAGATGAAAATTTGAAACTCAAAGATTACAGAAG
ATATCTGAAAATGGTTCGTTTCTGGGGCAGGGAAAAGGAAAATATAAAGCGGGAATTTGACAGTAAAAAATGGGAGA
GGTTTTTGCCAAGAGAACTCTGGCAGAAAAGAAACCTCGAAGATGCGTATCAACTGGCAAAAGAGAAAAACACCGAG
TTATTCAATAAATTGAAAACAACTGTTGAGAGAATGAACGAACTGGAATTCGAAAAGTATCAGCAGATAAACGACGC
AAAAGATTTGGCAAATTTAAGGCAACTGGCGCGGGACTTCGGCGTGAAGTGGGAAGAAAAGGACTGGCAAGAGTATT
CGGGGCAGATAAAAAAACAAATTACAGACAGGCAAAAACTTACAATAATGAAACAAAGGATTACTGCTGCATTGAAG
AAAAAGCAAGGCATAGAAAATCTTAATCTTAGGATAACAACCGACACCAATAAAAGCAGAAAGGTGGTATTGAACAG
AATAGCGCTACCTAAAGGTTTTGTAAGGAAGCATATCTTAAAAACAGATATAAAGATATCAAAGCAAATAAGGCAAT
CACAATGTCCTATTATACTGTCAAACAATTATATGAAGCTGGCAAAGGAATTCTTTGAGGAGAGAAATTTTGATAAG
ATGACGCAGATAAACGGGCTATTTGAGAAAAATGTACTTATAGCGTTTATGATAGTTTATCTGATGGAACAACTGAA
TCTTCGACTTGGTAAGAATACGGAACTTAGCAATCTTAAAAAAACGGAGGTTAATTTTACGATAACCGACAAGGTAA
CGGAAAAAGTCCAGATTTCGCAGTATCCATCGCTTGTTTTCGCCATAAACAGAGAATATGTTGATGGAATCAGCGGT
TATAAGTTACCGCCCAAAAAACCGAAAGAGCCTCCGTATACTTTCTTCGAGAAAATAGACGCAATAGAAAAAGAACG
AATGGAATTCATAAAACAGGTCCTCGGTTTCGAAGAACATCTTTTTGAGAAGAATGTAATAGACAAAACTCGCTTTA
CTGATACTGCGACTCATATAAGTTTTAATGAAATATGTGATGAGCTTATAAAAAAAGGATGGGACGAAAACAAAATA
ATAAAACTTAAAGATGCGAGGAATGCAGCATTGCATGGTAAGATACCGGAGGATACGTCTTTTGATGAAGCGAAAGT
ACTGATAAATGAATTAAAAAAATGA (SEQ ID NO: 21)

Human codon-optimized coding sequences for the seven Cas13e and Cas13f proteins (*i.e.*, Cas13e.1, Cas13e.2, Cas13f.1, Cas13f.2, Cas13f.3, Cas13f.4 and Cas13f.5), generated for further functional experiments, are SEQ ID NOs: 22-28, respectively.

ATGGCCCAGGTGAGCAAGCAGACCTCCAAGAAGAGGGAGCTGAGCATCGACGAGTACCAGGGCGCCCGGAAGTGGTG
CTTCACCATTGCCTTCAACAAGGCCCTGGTGAACCGGGACAAGAACGACGGCCTGTTCGTGGAAAGCCTGCTGAGAC
ACGAGAAGTACAGCAAGCACGACTGGTACGACGAAGATACCCGGGCCCTGATCAAGTGCAGCACCCAGGCCGCCAAC
GCCAAGGCTGAAGCCCTGCGGAACTACTTCAGTCACTACCGGCATAGCCCTGGCTGCCTGACCTTCACCGCCGAGGA
CGAACTGCGGACCATCATGGAGAGAGCCTATGAGCGGGCCATCTTCGAGTGCAGAAGAAGAGAGACAGAGGTGATCA
TCGAGTTTCCCAGCCTGTTCGAGGGCGACCGGATCACCACCGCCGGCGTGGTGTTTTTCGTGAGCTTTTTCGTGGAA
AGAAGAGTGCTGGATCGGCTGTATGGAGCCGTGTCCGGCCTGAAGAAGAATGAGGGACAGTACAAGCTGACCCGGAA
GGCCCTGAGCATGTACTGCCTGAAGGACAGCAGATTCACCAAGGCCTGGGATAAGCGGGTGCTGCTGTTCAGAGACA
TCCTGGCCCAGCTGGGAAGAATCCCCGCCGAGGCCTACGAGTACTACCACGGCGAGCAGGGTGATAAGAAGAGAGCT
AACGACAATGAGGGCACAAATCCCAAGCGGCACAAGGACAAGTTCATCGAATTTGCACTGCACTACCTGGAAGCCCA

GCACAGCGAGATCTGCTTCGGCAGACGCCACATCGTGCGGGAAGAGGCCGGCGCCGGCGATGAGCACAAGAAGCACC
GGACCAAGGGAAAGGTGGTGGTGGACTTCAGCAAGAAGGACGAGGACCAGAGCTACTATATCTCCAAGAACAACGTG
ATCGTGCGGATCGACAAGAACGCCGGCCCTAGAAGCTACCGGATGGGCCTGAACGAGCTGAAGTACCTCGTGCTGCT
GAGCCTGCAGGGGAAGGGCGACGATGCCATCGCCAAGCTGTACAGATACAGACAGCACGTGGAGAACATCCTGGATG
TGGTGAAGGTGACCGATAAGGATAACCACGTGTTCCTGCCCCGCTTCGTGCTGGAGCAGCACGGCATCGGCAGAAAG
GCCTTCAAGCAGCGGATCGATGGACGGGTGAAGCACGTGCGGGGCGTGTGGGAGAAGAAGAAGGCCGCCACCAATGA
AATGACCCTGCACGAGAAGGCCAGAGACATCCTGCAGTACGTGAACGAAAACTGCACCCGGTCCTTCAACCCTGGCG
AATACAACAGACTGCTGGTGTGCCTGGTGGGCAAGGACGTGGAGAACTTTCAGGCCGGCCTGAAGCGGCTGCAGCTG
GCCGAAAGGATCGATGGCCGGGTGTACTCCATCTTCGCCCAGACCAGCACCATCAATGAGATGCACCAGGTGGTGTG
CGACCAGATCCTGAACCGGCTGTGCAGAATCGGCGACCAGAAGCTGTACGATTACGTGGGACTGGGCAAGAAGGACG
AAATCGACTACAAGCAGAAGGTGGCCTGGTTCAAGGAGCACATCAGCATCCGGAGAGGATTCCTGAGAAAGAAGTTC
TGGTACGATAGCAAGAAGGGATTCGCAAAGCTGGTGGAGGAACACCTGGAGTCCGGCGGCGGCCAGCGCGACGTGGG
CCTGGACAAGAAGTACTACCACATCGACGCCATCGGCAGATTCGAGGGCGCCAACCCCGCCCTGTACGAGACCCTGG
CCAGAGATCGGCTGTGCCTCATGATGGCCCAGTACTTCCTGGGCAGCGTGAGAAAGGAACTGGGCAACAAGATTGTG
TGGAGCAACGACAGCATCGAACTGCCTGTGGAAGGCTCTGTGGGAAATGAGAAGAGCATCGTGTTCTCCGTGTCTGA
CTACGGCAAGCTGTACGTGCTGGACGATGCCGAATTCCTGGGCCGGATCTGCGAATACTTCATGCCCCACGAAAAGG
GCAAGATCCGGTACCACACAGTGTACGAAAAGGGCTTTAGAGCATACAACGACCTGCAGAAGAAGTGCGTGGAGGCC
GTGCTGGCTTTCGAAGAGAAGGTGGTGAAGGCCAAGAAGATGAGCGAGAAGGAAGGCGCCCACTACATCGACTTCCG
GGAGATCCTGGCCCAGACCATGTGCAAGGAGGCCGAGAAGACCGCAGTGAACAAGGTGAGACGCGCCTTCTTCCACC
ACCACCTGAAGTTCGTGATTGACGAGTTCGGCCTGTTCAGCGACGTGATGAAGAAGTACGGCATCGAGAAGGAATGG
AAGTTCCCTGTCAAGTAA (SEQ ID NO: 22)

ATGAAGGTGGAGAACATCAAGGAAAAGTCCAAGAAGGCTATGTATCTGATCAACCACTATGAAGGCCCTAAGAAGTG
GTGCTTCGCCATCGTGCTGAATAGGGCCTGCGACAACTATGAGGATAACCCCCACCTGTTCAGCAAGAGCCTGCTGG
AATTTGAAAAGACCAGCAGAAAGGACTGGTTCGACGAGGAGACCAGGGAACTGGTGGAGCAGGCCGACACCGAGATC
CAGCCCAACCCCAACCTGAAGCCTAACACCACCGCCAACAGAAAGCTGAAGGACATCCGGAACTACTTCAGCCACCA
CTACCACAAGAATGAGTGCCTGTACTTCAAGAACGACGACCCTATCCGGTGCATCATGGAGGCAGCCTACGAGAAGT
CCAAGATCTACATCAAGGGCAAGCAGATTGAGCAGTCCGACATCCCCCTCCCTGAGCTGTTTGAGTCTAGCGGCTGG
ATCACCCCAGCCGGCATCCTGCTGCTGGCCAGCTTCTTTGTGGAGAGAGGCATTCTGCACAGACTGATGGGCAACAT
CGGCGGCTTCAAGGACAACCGGGGCGAATACGGACTGACCCACGATATCTTCACCACCTACTGCCTGAAGGGCAGCT
ACTCCATCAGAGCCCAGGACCACGACGCCGTGATGTTCAGAGACATCCTGGGCTACCTGAGCAGAGTGCCGACCGAG
AGCTTTCAGCGCATCAAGCAGCCACAGATCAGAAAGGAGGGGCAGCTGAGCGAGCGGAAGACAGACAAGTTTATCAC
CTTCGCCCTGAACTACCTGGAAGATTATGGACTGAAGGATCTGGAAGGCTGCAAGGCCTGCTTCGCCCGGAGCAAGA
TCGTGAGAGAGCAGGAGAACGTGGAAAGCATCAATGACAAGGAGTACAAGCCTCACGAAAACAAGAAGAAGGTGGAA
ATCCACTTCGATCAGTCTAAGGAAGACCGGTTCTACATCAACCGGAACAACGTGATCCTGAAGATCCAGAAGAAGGA
CGGCCACAGCAACATCGTGAGAATGGGCGTGTACGAGCTGAAGTATCTGGTGCTGATGTCCCTGGTGGGCAAGGCCA
AGGAAGCCGTGGAGAAGATCGACAACTACATCCAGGATCTGAGAGACCAGCTGCCCTACATCGAGGGCAAGAACAAG
GAAGAAATCAAGGAGTACGTGAGATTCTTCCCCAGATTCATCAGATCCCACCTGGGCCTGCTGCAGATTAACGATGA
GGAGAAGATCAAGGCCCGGCTGGACTATGTGAAGACAAAGTGGCTGGACAAGAAGGAGAAGTCCAAGGAGCTGGAGC
TGCACAAGAAGGGCCGGGATATCCTGCGGTACATCAACGAGCGGTGCGACCGGGAGCTGAACCGGAACGTGTACAAC
CGGATCCTGGAGCTGCTGGTGAGCAAGGACCTGACCGGCTTCTACCGGGAGCTGGAGGAGCTGAAGCGGACCAGACG
GATCGATAAGAACATTGTGCAGAACCTGTCCGGCCAGAAGACCATCAACGCCCTGCACGAAAAGGTGTGCGATCTCG
TGCTGAAGGAGATCGAGAGCCTGGACACCGAGAACCTGCGGAAGTACCTGGGCCTGATCCCCAAGGAGGAGAAGGAA
GTGACCTTTAAGGAGAAGGTGGACAGGATCCTGAAGCAGCCGGTGATCTACAAGGGCTTCCTGCGGTACCAGTTCTT
CAAGGACGACAAGAAGAGCTTCGTGCTGCTGGTGGAAGACGCCCTGAAGGAGAAGGGAGGCGGCTGCGACGTGCCCC
TGGGCAAGGAGTACTACAAGATCGTGTCCCTGGACAAGTATGACAAGGAAAATAAGACCCTGTGCGAGACCCTGGCA
ATGGATAGACTGTGCCTGATGATGGCCCGGCAGTATTACCTGAGCCTGAACGCCAAGCTGGCCCAGGAGGCCCAGCA
GATCGAATGGAAGAAGGAGGATAGCATTGAGCTGATCATCTTCACACTGAAGAATCCTGACCAGTCCAAGCAGAGCT
TCTCCATCCGGTTCAGCGTGCGGGACTTCACCAAGCTGTACGTGACCGACGACCCCGAATTCCTGGCCCGGCTGTGC
AGCTACTTCTTCCCCGTGGAGAAGGAGATCGAATACCACAAGCTGTACTCTGAAGGCATTAACAAGTACACCAACCT
GCAGAAGGAGGGGATCGAAGCCATCCTGGAGCTGGAGAAGAAGCTGATCGAAAGAAACCGGATCCAGTCCGCCAAGA
ACTACCTGAGCTTTAACGAAATCATGAACAAGAGCGGCTACAACAAGGATGAGCAGGATGACCTGAAGAAGGTGAGG
AACTCCCTGCTGCACTACAAGCTGATCTTCGAAAAGGAGCACCTGAAGAAGTTCTATGAAGTGATGCGGGGCGAGGG
AATCGAGAAGAAGTGGTCCCTGATCGTGTAA (SEQ ID NO: 23)

ATGAATGGCATCGAGCTGAAGAAGGAAGAAGCCGCCTTCTACTTCAATCAGGCCGAGCTGAACCTGAAGGCCATTGA
GGACAACATCTTCGACAAGGAGAGACGGAAGACACTGCTGAACAACCCCCAGATCCTGGCCAAGATGGAGAACTTTA
TCTTCAATTTCCGGGACGTGACCAAGAACGCCAAGGGCGAAATCGACTGCCTGCTGCTGAAGCTGAGAGAGCTGCGG

AACTTTTACAGCCACTACGTGCACAAGCGGGACGTCAGAGAACTGAGCAAGGGCGAGAAGCCGATCCTGGAGAAGTA
CTACCAGTTCGCCATCGAATCCACCGGCTCTGAGAACGTGAAGCTCGAAATCATCGAAAACGACGCCTGGCTGGCCG
ACGCCGGCGTGCTGTTCTTCCTGTGCATCTTCCTGAAGAAGAGCCAGGCAAACAAGCTGATCAGCGGCATCAGCGGC
TTCAAGAGAAACGACGACACCGGCCAGCCTCGGAGAAACCTGTTCACCTACTTCTCCATCCGGGAGGGCTACAAGGT
GGTGCCCGAAATGCAGAAGCACTTCCTGCTGTTCTCCCTGGTGAACCACCTGAGCAACCAGGACGATTATATCGAAA
AGGCCCACCAGCCCTACGACATCGGCGAGGGCCTCTTCTTCCACCGGATTGCCAGCACCTTCCTGAACATCTCCGGA
ATCCTGAGAAACATGAAGTTCTACACCTATCAGAGCAAGAGACTGGTGGAGCAGAGAGGCGAGCTGAAGCGGGAAAA
GGACATCTTCGCCTGGGAAGAACCGTTTCAGGGCAATTCCTACTTTGAGATCAACGGCCACAAGGGCGTGATTGGCG
AAGACGAGCTGAAGGAGCTGTGCTACGCCTTCCTGATCGGCAACCAGGACGCCAACAAGGTGGAGGGCCGGATCACC
CAGTTCCTGGAGAAGTTCAGAAACGCCAACAGCGTGCAGCAGGTGAAGGACGACGAGATGCTGAAGCCTGAATATTT
CCCCGCCAACTACTTTGCCGAGAGCGGCGTGGGCCGGATCAAGGACCGGGTGCTGAACAGACTGAACAAGGCCATCA
AGAGCAACAAGGCCAAGAAGGGCGAGATCATCGCCTATGACAAGATGAGAGAAGTGATGGCTTTCATCAATAACTCT
CTGCCCGTGGACGAGAAGCTGAAGCCCAAGGATTACAAGAGATACCTGGGCATGGTGAGATTCTGGGATAGAGAAAA
GGACAATATCAAGCGCGAGTTCGAAACGAAGGAGTGGAGCAAGTATCTGCCCTCCAACTTCTGGACCGCCAAGAACC
TGGAGAGAGTGTACGGACTGGCCCGGGAAAAGAACGCAGAGCTGTTTAACAAGCTGAAGGCCGACGTGGAGAAGATG
GACGAAAGAGAGCTGGAAAAGTATCAGAAGATCAACGACGCCAAGGATCTGGCCAACCTGCGGCGGCTGGCCAGCGA
CTTCGGAGTGAAGTGGGAGGAGAAGGATTGGGACGAGTACTCCGGCCAGATCAAGAAGCAGATCACAGATTCCCAGA
AGCTGACCATCATGAAGCAGAGAATCACAGCCGGCCTGAAGAAGAAGCACGGCATCGAAAACCTGAACCTGAGGATC
ACCATCGACATCAACAAGTCCAGAAAGGCCGTGCTGAATCGGATCGCCATCCCCAGAGGATTTGTGAAGCGGCACAT
CCTGGGCTGGCAGGAATCCGAGAAGGTGAGCAAGAAGATCAGAGAAGCCGAATGCGAGATTCTGCTGAGCAAGGAGT
ACGAGGAGCTGAGCAAGCAGTTCTTTCAGAGCAAGGACTACGACAAGATGACCCGCATCAACGGCCTGTACGAGAAG
AATAAGCTGATCGCCCTGATGGCCGTGTATCTGATGGGGCAGCTGAGAATCCTGTTCAAGGAGCACACCAAGCTGGA
CGACATCACCAAGACCACCGTGGATTTCAAGATCAGCGACAAGGTGACCGTGAAGATCCCCTTCTCCAACTATCCCT
CCCTGGTGTACACCATGAGCAGCAAGTACGTGGACAATATCGGCAACTACGGCTTCAGCAACAAGGACAAGGATAAG
CCCATTCTGGGCAAGATCGACGTGATCGAGAAGCAGCGGATGGAGTTTATCAAGGAGGTGCTGGGATTCGAGAAGTA
CCTGTTTGACGATAAGATCATCGACAAGAGCAAGTTCGCCGACACCGCCACCCACATCAGCTTTGCCGAAATCGTGG
AAGAACTGGTGGAGAAGGGCTGGGACAAGGACCGGCTGACGAAGCTGAAGGATGCCCGGAACAAGGCCCTGCACGGC
GAGATCCTGACCGGCACCAGCTTCGACGAGACAAAGTCCCTGATCAACGAGCTGAAGAAGTAA (SEQ ID NO:
24)

ATGAGCCCTGATTTCATCAAGCTGGAGAAGCAGGAAGCAGCCTTCTACTTTAACCAGACCGAGCTGAACCTGAAGGC
CATCGAATCCAATATCCTGGATAAGCAGCAGAGAATGATCCTGCTGAACAACCCCAGAATCCTGGCCAAGGTGGGCA
ACTTCATCTTCAATTTCCGGGACGTGACCAAGAACGCAAAGGGCGAAATCGACTGCCTGCTGTTCAAGCTGGAGGAA
CTGCGGAACTTCTACAGCCACTACGTGCACACCGATAACGTGAAGGAACTGTCCAACGGAGAGAAGCCTCTGCTGGA
GCGGTACTACCAGATCGCCATCCAGGCCACAAGAAGCGAGGACGTGAAGTTCGAGCTGTTCGAGACCAGGAACGAGA
ACAAGATCACCGACGCAGGCGTGCTGTTCTTCCTGTGCATGTTCCTGAAGAAGAGCCAGGCTAATAAGCTGATTTCC
GGCATCAGCGGCTTCAAGCGGAACGACCCCACCGGCCAGCCCAGACGGAACCTCTTTACCTACTTCTCTGCCCGGGA
GGGCTACAAGGCCCTGCCTGACATGCAGAAGCACTTCCTGCTGTTCACCCTGGTGAACTACCTGAGCAACCAGGACG
AGTACATCTCCGAGCTGAAGCAGTACGGAGAGATCGGACAGGGAGCCTTCTTCAACAGAATCGCCAGCACCTTCCTG
AACATCAGCGGCATCAGCGGCAACACCAAGTTCTACAGCTACCAGAGCAAGAGAATCAAGGAGCAGCGGGGCGAACT
GAACAGCGAAAAGGACAGCTTCGAGTGGATCGAGCCCTTTCAGGGCAACTCTTATTTTGAGATCAACGGCCACAAGG
GCGTGATCGGCGAAGACGAGCTGAAGGAGCTGTGCTACGCCCTGCTGGTGGCCAAGCAGGACATCAATGCCGTGGAG
GGAAAGATCATGCAGTTCCTGAAGAAGTTCAGGAACACCGGCAACCTGCAGCAGGTGAAGGACGACGAGATGCTGGA
AATCGAGTACTTTCCCGCCAGCTACTTCAACGAGAGCAAGAAGGAGGACATCAAGAAGGAGATCCTGGGCAGACTGG
ACAAGAAGATCCGGTCCTGCAGCGCCAAGGCCGAGAAGGCCTACGACAAGATGAAGGAGGTGATGGAGTTTATCAAT
AACAGCCTGCCCGCCGAGGAGAAGCTGAAGAGGAAGGACTACCGCAGATACCTGAAGATGGTGAGATTCTGGTCCAG
AGAAAAGGGCAACATCGAGAGAGAGTTCAGAACCAAGGAGTGGTCCAAGTACTTCAGCAGCGACTTCTGGAGAAAGA
ACAATCTGGAGGATGTGTACAAGCTGGCCACCCAGAAGAACGCCGAGCTGTTCAAGAATCTGAAGGCCGCCGCCGAG
AAGATGGGCGAAACAGAATTCGAAAAGTACCAGCAGATCAACGATGTGAAGGACCTGGCCAGCCTGAGACGGCTGAC
CCAGGATTTCGGCCTGAAGTGGGAGGAGAAGGATTGGGAGGAGTACAGCGAACAGATCAAGAAGCAGATCACCGACC
GGCAGAAGCTGACAATCATGAAGCAGCGGGTGACCGCCGAGCTGAAGAAGAAGCACGGCATCGAGAATCTGAACCTC
AGAATTACCATCGATTCCAACAAGAGCAGAAAGGCCGTGCTGAACAGAATCGCCATTCCCCGGGGCTTCGTGAAGAA
GCACATTCTGGGCTGGCAGGGCAGCGAAAAGATCAGCAAGAATATCCGGGAGGCCGAGTGCAAGATCCTGCTGTCCA
AGAAGTATGAGGAGCTGTCTCGGCAGTTCTTTGAGGCTGGCAACTTCGACAAGCTGACCCAGATCAACGGCCTGTAC
GAAAAGAATAAGCTGACCGCCTTCATGTCCGTCTACCTGATGGGCAGACTGAACATCCAGCTGAACAAGCACACGGA
GCTGGGAAATCTGAAGAAGACCGAGGTGGACTTCAAGATTTCCGACAAGGTGACAGAAAAGATCCCCTTCTCCCAGT
ACCCTAGCCTGGTGTACGCTATGAGCCGGAAGTACGTGGACAACGTGGACAAGTACAAGTTCAGCCACCAGGACAAG

AAGAAGCCCTTCCTGGGCAAGATCGACAGCATCGAAAAGGAGAGAATCGAATTCATCAAGGAGGTGCTGGACTTCGA
AGAGTACCTGTTTAAGAACAAGGTGATCGACAAGAGCAAGTTCAGCGATACCGCCACCCATATCTCTTTCAAGGAAA
TCTGCGACGAGATGGGCAAGAAGGGCTGCAACCGCAACAAGCTGACCGAGCTGAATAACGCTAGAAACGCCGCACTG
CACGGAGAAATCCCCAGCGAGACCAGCTTCCGGGAGGCCAAGCCCCTGATCAACGAACTGAAGAAGTAA (SEQ ID
NO: 25)

ATGAGCCCTGACTTCATCAAGCTGGAAAAGCAGGAAGCCGCCTTCTACTTTAATCAGACCGAGCTGAACCTGAAGGC
CATCGAGAGCAACATCTTCGACAAGCAGCAGCGGGTGATCCTGCTGAATAACCCCCAGATCCTGGCCAAGGTGGGCG
ACTTCATCTTCAACTTCCGGGACGTGACCAAGAACGCCAAGGGAGAAATCGACTGCCTGCTGCTGAAGCTGCGGGAG
CTGAGAAACTTCTACAGCCACTATGTGTACACCGACGACGTGAAGATCCTGAGCAACGGCGAGAGGCCCCTGCTGGA
GAAGTACTACCAGTTTGCCATCGAGGCCACCGGATCTGAGAATGTGAAGCTGGAGATCATCGAGAGCAACAACCGGC
TGACCGAAGCGGGCGTGCTGTTCTTCCTGTGCATGTTCCTGAAGAAGAGCCAGGCCAACAAGCTGATTTCCGGCATC
TCCGGATTCAAGCGCAACGACCCTACCGGACAGCCTCGGCGGAACCTGTTCACCTACTTTAGCGTGCGGGAGGGCTA
CAAGGTGGTGCCCGACATGCAGAAGCACTTCCTGCTGTTCGTGCTGGTGAACCACCTGTCCGGCCAGGATGACTATA
TTGAGAAGGCCCAGAAGCCCTACGACATCGGCGAAGGCCTGTTCTTCCACAGAATCGCCAGCACCTTTCTCAACATC
AGCGGCATCCTGAGAAACATGGAATTCTACATCTACCAGAGCAAGCGGCTGAAGGAGCAGCAGGGAGAGCTGAAGAG
AGAGAAGGACATCTTCCCTTGGATCGAGCCTTTCCAGGGCAACAGCTACTTTGAGATCAACGGAAACAAGGGCATCA
TCGGCGAGGACGAACTGAAGGAACTGTGCTACGCCCTGCTGGTGGCCGGCAAGGACGTGAGAGCCGTGGAAGGAAAG
ATCACCCAGTTCCTGGAGAAGTTCAAGAACGCCGATAACGCCCAGCAGGTGGAGAAGGATGAAATGCTGGACCGGAA
CAACTTCCCTGCCAATTACTTTGCCGAAAGCAACATCGGCAGCATCAAGGAAAAGATCCTGAATAGACTGGGCAAGA
CCGACGACTCCTACAACAAGACCGGCACCAAGATCAAGCCCTACGACATGATGAAGGAGGTGATGGAGTTCATCAAT
AATTCTCTGCCCGCCGATGAGAAGCTGAAGCGGAAGGACTACCGGAGATACCTGAAGATGGTCCGGATCTGGGACAG
CGAAAAGGACAATATCAAGCGGGAGTTTGAGAGCAAGGAATGGAGCAAGTATTTCAGCAGCGACTTCTGGATGGCCA
AGAACCTGGAAAGAGTGTACGGCCTGGCCAGGGAAAAGAACGCCGAGCTGTTTAACAAGCTGAAGGCCGTGGTGGAG
AAGATGGACGAGCGGGAGTTCGAAAAGTACCGGCTGATCAACAGCGCCGAAGACCTGGCCAGCCTGCGGAGACTGGC
CAAGGACTTCGGCCTGAAGTGGGAGGAGAAGGACTGGCAGGAGTATTCTGGCCAGATCAAGAAGCAGATCTCCGACA
GACAGAAGCTGACAATTATGAAGCAGCGGATCACAGCCGAACTGAAGAAGAAGCACGGAATCGAGAACCTGAATCTG
CGGATCACCATCGACAGCAACAAGTCCAGAAAGGCCGTGCTGAACCGGATCGCCGTGCCCCGGGGCTTCGTGAAGGA
ACACATCCTGGGCTGGCAAGGCTCTGAAAAGGTGAGCAAGAAGACCAGAGAAGCCAAGTGCAAGATCCTGCTGAGCA
AGGAGTACGAGGAACTGAGCAAGCAGTTCTTTCAGACACGGAATTACGACAAGATGACCCAGGTGAACGGCCTGTAC
GAGAAGAACAAGCTGCTGGCCTTCATGGTGGTGTACCTGATGGAGAGACTGAACATCCTGCTGAACAAGCCCACAGA
GCTGAACGAACTGGAAAAGGCCGAAGTGGACTTCAAGATCTCCGACAAGGTGATGGCCAAGATCCCTTTCTCTCAGT
ACCCCAGCCTGGTGTATGCAATGAGCTCCAAGTACGCCGACAGCGTGGGCTCTTACAAGTTCGAAAACGACGAGAAG
AACAAGCCCTTTCTGGGCAAGATCGACACAATCGAGAAGCAGAGAATGGAGTTCATCAAGGAGGTGCTGGGCTTCGA
GGAATACCTGTTCGAGAAGAAGATCATCGATAAGAGCGAATTCGCCGACACCGCCACCCACATCAGCTTCGACGAGA
TCTGCAACGAGCTGATCAAGAAGGGCTGGGACAAGGACAAGCTGACCAAGCTGAAGGACGCCCGGAACGCCGCCCTG
CACGGCGAGATCCCCGCCGAGACCAGCTTCCGGGAGGCCAAGCCCCTGATTAACGGCCTGAAGAAGTAA (SEQ ID
NO: 26)

ATGAACATCATCAAGCTGAAGAAGGAGGAAGCCGCCTTTTACTTTAACCAGACAATCCTGAATCTGAGCGGCCTGGA
CGAGATCATCGAGAAGCAGATCCCCCACATCATCTCCAATAAGGAAAACGCCAAGAAGGTGATTGATAAGATCTTCA
ATAACAGACTGCTGCTGAAGAGCGTGGAAAACTATATCTACAACTTCAAGGACGTGGCCAAGAACGCCCGGACCGAA
ATCGAAGCCATCCTGCTGAAGCTGGTGGAGCTGAGAAACTTCTACTCCCACTACGTGCACAACGACACCGTGAAGAT
CCTGTCCAATGGCGAGAAGCCCATCCTGGAAAAGTACTACCAGATCGCCATCGAAGCCACCGGCTCTAAGAACGTGA
AGCTGGTCATTATCGAAAACAACAACTGCCTGACCGACTCCGGCGTGCTGTTCCTGCTGTGCATGTTCCTGAAGAAG
AGCCAGGCCAACAAGCTGATTAGCAGCGTGAGCGGCTTTAAGCGGAACGACAAGGAAGGCCAGCCCAGAAGGAACCT
CTTTACTTACTATAGCGTGAGGGAAGGCTACAAGGTGGTGCCAGACATGCAGAAGCACTTCCTGCTGTTCGCCCTGG
TCAACCACCTGTCCGAGCAGGACGACCACATCGAGAAGCAGCAGCAGAGCGACGAGCTGGGCAAGGGCCTGTTCTTC
CACAGAATCGCCAGCACATTCCTGAATGAAAGCGGCATCTTCAACAAGATGCAGTTTTACACCTACCAGAGCAATCG
GCTGAAGGAGAAGCGGGGCGAGCTGAAGCACGAGAAGGACACCTTCACCTGGATCGAGCCTTTCCAGGGAAACAGCT
ACTTCACCCTGAACGGGCACAAGGGCGTGATCAGCGAGGATCAGCTGAAGGAACTGTGCTACACAATCCTGATCGAG
AAGCAGAACGTGGACAGCCTGGAGGGCAAGATCATTCAGTTCCTGAAGAAGTTTCAGAACGTGTCTAGCAAGCAGCA
GGTGGATGAGGACGAGCTGCTGAAGCGGGAATACTTCCCCGCCAACTACTTCGGCCGGGCCGGCACCGGCACCCTGA
AGGAGAAGATCCTGAACCGGCTGGACAAGCGGATGGACCCCACCAGCAAGGTGACCGACAAGGCCTATGACAAGATG
ATCGAGGTGATGGAGTTCATCAACATGTGCCTGCCCAGCGACGAGAAGCTGCGGCAGAAGGATTACCGGAGATATCT
GAAGATGGTCAGATTCTGGAACAAGGAGAAGCACAACATCAAGAGAGAATTCGACAGCAAGAAGTGGACCAGATTCC
TGCCCACCGAGCTGTGGAATAAGCGGAACCTGGAGGAAGCCTACCAGCTGGCCCGGAAGGAGAACAAGAAGAAGCTG
GAGGACATGAGGAATCAGGTGAGGAGCCTGAAGGAGAACGACCTGGAGAAGTACCAGCAGATCAACTATGTGAACGA

CCTGGAAAACCTGCGGCTGCTGTCCCAAGAGCTGGGCGTGAAGTGGCAGGAGAAGGACTGGGTGGAATACAGCGGCC
AGATCAAGAAGCAGATCAGCGATAACCAGAAGCTGACAATCATGAAGCAGAGAATCACCGCCGAGCTGAAGAAGATG
CACGGCATCGAGAACCTGAACCTGAGAATCAGCATCGACACCAACAAGTCCCGGCAGACTGTGATGAACAGAATTGC
CCTGCCCAAGGGCTTCGTGAAGAACCACATTCAGCAGAACAGCAGCGAGAAGATCAGCAAGAGAATCAGAGAGGACT
ACTGCAAGATCGAGCTGTCCGGCAAGTACGAAGAGCTGAGCAGACAGTTTTTCGACAAGAAGAACTTTGACAAGATG
ACCCTGATCAACGGACTGTGCGAGAAGAATAAGCTCATCGCCTTCATGGTGATTTACCTGCTGGAGCGGCTGGGCTT
CGAGCTGAAGGAGAAGACCAAGCTGGGCGAGCTGAAGCAGACCCGGATGACATATAAGATCAGCGACAAGGTGAAGG
AGGACATCCCCCTCTCCTACTACCCCAAGCTGGTGTACGCCATGAATCGGAAGTATGTGGACAACATCGATAGCTAC
GCCTTCGCCGCCTACGAGTCTAAGAAGGCCATCCTGGACAAGGTGGACATCATTGAGAAGCAGAGAATGGAATTCAT
CAAGCAGGTGCTGTGCTTCGAGGAATACATCTTCGAGAACAGAATCATCGAGAAGAGCAAGTTCAACGATGAGGAGA
CCCACATCAGCTTCACCCAGATCCACGACGAACTGATCAAGAAGGGCAGAGATACCGAAAAGCTGAGCAAGCTGAAG
CACGCCAGAAACAAGGCCCTGCACGGCGAGATCCCCGACGGGACCAGCTTTGAGAAGGCCAAGCTGCTGATCAACGA
AATCAAGAAGTAA (SEQ ID NO: 27)

ATGAACGCCATCGAGCTGAAGAAGGAAGAGGCCGCCTTCTACTTCAACCAGGCCAGACTGAACATCTCTGGCCTGGA
CGAAATCATCGAGAAGCAACTGCCACACATCGGCTCTAACAGAGAGAACGCCAAGAAGACTGTGGACATGATCCTGG
ATAACCCCGAGGTGCTGAAGAAGATGGAAAACTACGTGTTCAACTCCCGCGATATTGCCAAGAATGCCCGGGGCGAG
CTGGAGGCCCTGCTGCTGAAGCTGGTCGAGCTGAGAAACTTCTATAGCCACTACGTGCACAAGGACGACGTCAAGAC
ACTGAGCTACGGTGAGAAGCCTCTGCTGGATAAGTACTACGAGATCGCCATCGAAGCCACCGGATCCAAGGACGTGC
GGCTGGAGATCATTGACGACAAGAATAAGCTGACCGACGCCGGAGTGCTGTTCCTGCTGTGCATGTTCCTGAAGAAG
AGCGAGGCTAACAAGCTGATTTCCAGCATCCGGGGCTTCAAGAGGAACGACAAGGAGGGCCAGCCTAGAAGAAACCT
GTTCACCTACTACAGCGTGAGAGAGGGCTATAAGGTGGTGCCCGACATGCAGAAGCACTTTCTGCTGTTCACCCTGG
TGAACCACCTGTCCAATCAGGACGAGTACATCTCCAACCTGCGCCCAAACCAGGAAATCGGCCAGGGCGGATTTTTC
CACCGGATCGCCAGCAAGTTCCTGAGCGACAGCGGAATCCTGCACAGCATGAAGTTCTACACATACAGATCCAAGCG
GCTGACCGAGCAGCGGGGAGAGCTGAAGCCCAAGAAGGACCACTTTACATGGATCGAGCCTTTCCAGGGCAATTCCT
ACTTCAGCGTGCAGGGCCAGAAGGGCGTGATCGGAGAGGAGCAGCTCAAGGAGCTGTGCTACGTGCTGCTGGTGGCC
CGGGAGGACTTCAGAGCCGTGGAGGGCAAGGTGACCCAGTTCCTGAAGAAGTTCCAGAATGCCAATAACGTGCAGCA
GGTGGAGAAGGACGAGGTGCTGGAAAAGGAGTACTTCCCCGCCAACTACTTTGAGAACCGGGACGTGGGAAGAGTCA
AGGACAAGATCCTGAACAGACTGAAGAAGATCACCGAGAGTTATAAGGCCAAGGGTAGAGAGGTGAAGGCCTACGAC
AAGATGAAGGAAGTGATGGAGTTCATCAACAACTGCCTGCCCACCGATGAAAACCTGAAGCTGAAGGACTACCGGCG
GTACCTGAAGATGGTGAGATTCTGGGGCAGAGAGAAGGAAAACATCAAGCGGGAGTTCGACTCCAAGAAGTGGGAGC
GCTTTCTCCCCCGGGAGCTGTGGCAGAAGAGAAACCTGGAGGACGCCTACCAGCTCGCCAAGGAGAAGAACACAGAG
CTGTTCAACAAGCTGAAGACCACCGTGGAGAGAATGAACGAACTGGAGTTCGAGAAGTACCAGCAGATCAATGACGC
CAAGGACCTGGCCAACCTGAGACAGCTGGCCAGAGACTTTGGAGTGAAGTGGGAGGAAAAGGACTGGCAGGAATACT
CTGGACAGATCAAGAAGCAGATCACCGACCGGCAGAAGCTGACCATCATGAAGCAGCGGATCACCGCCGCCCTGAAG
AAGAAGCAGGGAATCGAAAACCTGAACCTGAGAATCACAACAGATACGAATAAGAGCAGGAAGGTGGTGCTGAACCG
GATCGCACTGCCCAAGGGATTCGTCAGAAAGCACATCCTGAAGACCGACATCAAGATCAGCAAGCAGATCCGGCAGA
GCCAGTGCCCTATCATCCTGTCTAACAACTACATGAAGCTGGCCAAGGAGTTCTTTGAAGAGCGGAACTTCGATAAG
ATGACCCAGATCAATGGCCTGTTCGAGAAGAACGTGCTGATCGCCTTCATGATCGTGTACCTGATGGAGCAGCTGAA
CCTGAGACTGGGCAAGAACACCGAGCTGTCCAACCTGAAGAAGACCGAGGTGAACTTTACCATCACCGACAAGGTGA
CCGAGAAGGTGCAAATCTCCCAGTACCCCAGCCTGGTGTTCGCCATTAACCGGGAGTACGTGGACGGCATCAGCGGC
TACAAGCTGCCCCCCAAGAAGCCCAAGGAACCTCCCTACACCTTCTTCGAAAAGATCGACGCCATCGAAAAGGAGCG
GATGGAATTCATCAAGCAGGTGCTGGGCTTCGAGGAGCACCTCTTCGAAAAGAACGTGATCGACAAGACCCGGTTTA
CCGACACCGCCACCCACATCAGCTTCAATGAGATCTGCGATGAGCTGATCAAGAAGGGCTGGGACGAAAACAAGATC
ATCAAGCTGAAGGATGCACGGAACGCTGCCCTGCACGGCAAGATCCCTGAAGATACCTCCTTTGACGAAGCCAAGGT
GCTGATCAACGAACTGAAGAAGTAA (SEQ ID NO: 28)

The seven CRISPR/Cas13e and Cas13f loci structures were shown in FIG. 1.

Further analysis of RNA secondary structures for the seven DR sequences in the pre-crRNA was conducted using RNAfold. The results were shown in FIG. 2. It is apparent that all shared very conserved secondary structure.

For example, in the Cas13e family, each DR sequence forms a secondary structure consisting of a 4-base pair stem (5'-GCUG-3'), followed by a symmetrical bulge of 5 + 5

nucleotides (excluding the 4 stem nucleotides), further followed by a 5-base pair stem (5'-GCC C/U C-3'), and a terminal 8-base loop (5'-CGAUUUGU-3', excluding the 2 stem nucleotides).

Likewise, in the Cas13f family, with one exception (Cas13f.4), each DR sequence forms a secondary structure consisting of a 5-base pair stem (5'GCUGU3'), followed by a nearly symmetrical bulge of 5 + 4 nucleotides (excluding the 4 stem nucleotides), further followed by a 6-base pair stem (5'A/G CCUCG3'), and a terminal 5-base loop (5'AUUUG3', excluding the 2 stem nucleotides). The only exception being the DR for Cas13f.4, in which the second step is 1 base pair shorter, and 2 additional bases were added to the first bulge to form a largely symmetrical 6 + 5 bulge.

Multi-sequence alignment of Cas13e and Cas13f proteins and the previously identified Cas13a, Cas13b, Cas13c, and Cas13d family proteins, using MAFFT, revealed that Cas13e and Cas13f proteins are relatively closest to the Cas13b proteins on the phylogenetic tree (FIG. 3).

Further, in terms of the locations of the RXXXXH motifs with respect to the N- and C-termini of the Cas proteins, Cas13e and Cas13f proteins, and to a lesser extent Cas13b proteins, have their RXXXXH motifs closer to their N- and C-termini, as compared to the Cas13a, Cas13c, and Cas13d (see FIG. 4).

I-TASSER was then used to predict 3D structures for Cas13e proteins, followed by visualization of the predicted structures using PyMOL. Although the two RXXXXH motifs are located very close to the N- and C-termini of Cas13e.1, they are very close by in the 3D structure (FIG. 5).

Example 2 Cas13e is an Effector RNase

In order to confirm that the newly identified Cas13e proteins are effective RNase functioning in the CRISPR/Cas system, Cas13e.1 coding sequence was codon optimized for human expression (SEQ ID NO: 22), and cloned into a first plasmid with GFP gene. Meanwhile, coding sequence for guide RNA (gRNA) targeting the reporter gene (mCherry) mRNA was cloned into a second plasmid with GFP gene. The gRNA consists of a spacer coding region flanked by two direct repeat sequences for Cas13e.1 (SEQ ID NO: 29). The sequence of GFP and mCherry reporter genes are SEQ ID NO: 30-31, respectively.

GCTGGAGCAGCCCCCGATTTGTGGGGTGATTACAGCGGTCTTCGATATTCAAGCGTCGGAAGACCTGCTGGAGCAGC

CCCCGATTTGTGGGGTGATTACAGC (SEQ ID NO: 29)

ATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTC
CGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGA
AGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTAC
GTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAA
CTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGC
TGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGG
ATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGC
TGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACA
TCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATG
GACGAGCTGTACAAGTAA (SEQ ID NO: 30)

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCA
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCG
GCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGAC
CACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGA
CGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCA
TCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATG
GCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGC
CGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGT
CCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACT
CTCGGCATGGACGAGCTGTACAAGTGA (SEQ ID NO: 31)

HEK293T cells were cultured in 24-well tissue culture plates according to standard protocol, and were used for triple plasmid transfection using LIPOFECTAMINE® 3000 and P3000™ reagent to introduce the three plasmids encoding the Cas13e.1 protein, the mCherry-targeting gRNA, and the mCherry coding sequence, respectively. In a negative control experiment, instead of using the plasmid encoding the mCherry-targeting gRNA, a control plasmid encoding a non-Target-gRNA was used. A GFP coding sequence was present in the Cas13e.1 and gRNA plasmid, thus expression of GFP can be used as an internal control for transfection success / efficiency. See schematic illustration in FIG. 6. Transfected HEK293T cells were then incubated at 37°C under 5% CO_2 for about 24 hours, before the cells were subject to examination under the fluorescent microscope.

As shown in FIG. 7, cells transfected with the mCherry-targeting gRNA, and cells transfected with the control non-targeting (NT) gRNA had equivalent growth and morphology in bright field microscope, and GFP expression in both were largely equivalent. However, RFP signal from mCherry expression was dramatically reduced by up to 75% based on flow cytometry analysis (FIG. 8). This suggests that Cas13e can utilize the mCherry-targeting gRNA to efficiently knock down mCherry mRNA level, and consequently mCherry protein expression.

Example 3 Effective Direction of sgRNA for Cas13e

Since Cas13e system can in theory utilize either the DR + Spacer (5'DR) or the Spacer + DR (3'DR) orientation, this experiment was designed to determine which is the correct orientation utilized by Cas13e.

Using a similar triple transfection experiment setting as in Example 2, it was found that only the 3'DR orientation (Spacer + DR) supported significant mCherry knock down. This demonstrated that Cas13e utilizes its crRNA with the DR sequence at the 3'-end of the spacer. See FIG. 9.

SgRNA of DR + Spacer (5' DR) and Spacer + DR (3' DR) are SEQ ID NOs: 32 and 33, respectively.

GCTGGAGCAGCCCCCGATTTGTGGGGTGATTACAGCGGTCTTCGATATTCAAGCGTCGGAAGACCT (SEQ ID NO: 32)

GGTCTTCGATATTCAAGCGTCGGAAGACCTGCTGGAGCAGCCCCCGATTTGTGGGGTGATTACAGC (SEQ ID NO: 33)

Example 4 Effect of Spacer Sequence Length on Specific Activity and Collateral Activity of Cas13e.1

In order to study the effect of spacer sequence length on specific activity and collateral activity of Cas13e.1, a set of sgRNA targeting the mCherry reporter gene were designed, with spacer sequence length of 20 nt, 25 nt, 30 nt, 35 nt, 40 nt, 45 nt, or 50 nt (SEQ ID NO: 34-40).

TTGGTGCCGCGCAGCTTCAC (SEQ ID NO: 34)

TTGGTGCCGCGCAGCTTCACCTTGT (SEQ ID NO: 35)

TTGGTGCCGCGCAGCTTCACCTTGTAGATG (SEQ ID NO: 36)

TTGGTGCCGCGCAGCTTCACCTTGTAGATGAACTC (SEQ ID NO: 37)

TTGGTGCCGCGCAGCTTCACCTTGTAGATGAACTCGCCGT (SEQ ID NO: 38)

TTGGTGCCGCGCAGCTTCACCTTGTAGATGAACTCGCCGTCCTGC (SEQ ID NO: 39)

TTGGTGCCGCGCAGCTTCACCTTGTAGATGAACTCGCCGTCCTGCAGGGA (SEQ ID NO: 40)

Using a similar triple transfection experiment setting as in Example 2, the knock down efficiency of mCherry and GFP gene were analyzed by flow cytometry.

The results of mCherry and GFP knock down experiments showed the specific activity and non-specific activity (collateral activity) of Cas13e.1, respectively. It was found that

Cas13e.1 has high specific activity with spacer lengths between about 30 nt to about 50 nt. See FIG. 10. Meanwhile. Cas13e.1 has highest non-specific activity when spacer length is about 30 nt. See FIG. 11.

Example 5 Single-Base RNA Editing using dCas13e.1-ADAR2DD Fusion

In order to test whether Cas13e can be used for RNA single base editing, dCas13e.1 was generated by mutating the two RXXXXH motifs to eliminate RNase activity. Then a high fidelity ADAR2DD mutant with E488Q and T375G double mutation was fused to the (C-terminus) of dCas13e.1 to create a putative A-to-G single base RNA editor named dCas13e.1-ADAR2DD. See coding sequence in SEQ ID NO: 41.

ATGCCCAAGAAGAAGCGGAAGGTGGCCCAGGTGAGCAAGCAGACCTCCAAGAAGAGGGAGCTGAGCATCGACGAGTA
CCAGGGCGCCCGGAAGTGGTGCTTCACCATTGCCTTCAACAAGGCCCTGGTGAACCGGGACAAGAACGACGGCCTGT
TCGTGGAAAGCCTGCTGAGACACGAGAAGTACAGCAAGCACGACTGGTACGACGAAGATACCCGGGCCCTGATCAAG
TGCAGCACCCAGGCCGCCAACGCCAAGGCTGAAGCCCTGGCGAACTACTTCAGTGCTTACCGGCATAGCCCTGGCTG
CCTGACCTTCACCGCCGAGGACGAACTGCGGACCATCATGGAGAGAGCCTATGAGCGGGCCATCTTCGAGTGCAGAA
GAAGAGAGACAGAGGTGATCATCGAGTTTCCCAGCCTGTTCGAGGGCGACCGGATCACCACCGCCGGCGTGGTGTTT
TTCGTGAGCTTTTTCGTGGAAAGAAGAGTGCTGGATCGGCTGTATGGAGCCGTGTCCGGCCTGAAGAAGAATGAGGG
ACAGTACAAGCTGACCCGGAAGGCCCTGAGCATGTACTGCCTGAAGGACAGCAGATTCACCAAGGCCTGGGATAAGC
GGGTGCTGCTGTTCAGAGACATCCTGGCCCAGCTGGGAAGAATCCCCGCCGAGGCCTACGAGTACTACCACGGCGAG
CAGGGTGATAAGAAGAGAGCTAACGACAATGAGGGCACAAATCCCAAGCGGCACAAGGACAAGTTCATCGAATTTGC
ACTGCACTACCTGGAAGCCCAGCACAGCGAGATCTGCTTCGGCAGACGCCACATCGTGCGGGAAGAGGCCGGCGCCG
GCGATGAGCACAAGAAGCACCGGACCAAGGGAAAGGTGGTGGTGGACTTCAGCAAGAAGGACGAGGACCAGAGCTAC
TATATCTCCAAGAACAACGTGATCGTGCGGATCGACAAGAACGCCGGCCCTAGAAGCTACCGGATGGGCCTGAACGA
GCTGAAGTACCTCGTGCTGCTGAGCCTGCAGGGGAAGGGCGACGATGCCATCGCCAAGCTGTACAGATACAGACAGC
ACGTGGAGAACATCCTGGATGTGGTGAAGGTGACCGATAAGGATAACCACGTGTTCCTGCCCCGCTTCGTGCTGGAG
CAGCACGGCATCGGCAGAAAGGCCTTCAAGCAGCGGATCGATGGACGGGTGAAGCACGTGCGGGGCGTGTGGGAGAA
GAAGAAGGCCGCCACCAATGAAATGACCCTGCACGAGAAGGCCAGAGACATCCTGCAGTACGTGAACGAAAACTGCA
CCCGGTCCTTCAACCCTGGCGAATACAACAGACTGCTGGTGTGCCTGGTGGGCAAGGACGTGGAGAACTTTCAGGCC
GGCCTGAAGCGGCTGCAGCTGGCCGAAAGGATCGATGGCCGGGTGTACTCCATCTTCGCCCAGACCAGCACCATCAA
TGAGATGCACCAGGTGGTGTGCGACCAGATCCTGAACCGGCTGTGCAGAATCGGCGACCAGAAGCTGTACGATTACG
TGGGACTGGGCAAGAAGGACGAAATCGACTACAAGCAGAAGGTGGCCTGGTTCAAGGAGCACATCAGCATCCGGAGA
GGATTCCTGAGAAAGAAGTTCTGGTACGATAGCAAGAAGGGATTCGCAAAGCTGGTGGAGGAACACCTGGAGTCCGG
CGGCGGCCAGCGCGACGTGGGCCTGGACAAGAAGTACTACCACATCGACGCCATCGGCAGATTCGAGGGCGCCAACC
CCGCCCTGTACGAGACCCTGGCCAGAGATCGGCTGTGCCTCATGATGGCCCAGTACTTCCTGGGCAGCGTGAGAAAG
GAACTGGGCAACAAGATTGTGTGGAGCAACGACAGCATCGAACTGCCTGTGGAAGGCTCTGTGGGAAATGAGAAGAG
CATCGTGTTCTCCGTGTCTGACTACGGCAAGCTGTACGTGCTGGACGATGCCGAATTCCTGGGCCGGATCTGCGAAT
ACTTCATGCCCCACGAAAAGGGCAAGATCCGGTACCACACAGTGTACGAAAAGGGCTTTAGAGCATACAACGACCTG
CAGAAGAAGTGCGTGGAGGCCGTGCTGGCTTTCGAAGAGAAGGTGGTGAAGGCCAAGAAGATGAGCGAGAAGGAAGG
CGCCCACTACATCGACTTCCGGGAGATCCTGGCCCAGACCATGTGCAAGGAGGCCGAGAAGACCGCAGTGAACAAGG
TGGCGGCTGCCTTCTTCGCTGCGCACCTGAAGTTCGTGATTGACGAGTTCGGCCTGTTCAGCGACGTGATGAAGAAG
TACGGCATCGAGAAGGAATGGAAGTTCCCTGTCAAGCCCAAGAAGAAGCGGAAGGTGGGTGGAGGCGGAGGTTCTGG
GGGAGGAGGTAGTGGCGGTGGTGGTTCAGGAGGCGGCGGAAGCCAGCTGCATTTACCGCAGGTTTTAGCTGACGCTG
TCTCACGCCTGGTCCTGGGTAAGTTTGGTGACCTGACCGACAACTTCTCCTCCCCTCACGCTCGCAGAAAAGTGCTG
GCTGGAGTCGTCATGACAACAGGCACAGATGTTAAAGATGCCAAGGTGATAAGTGTTTCTACAGGAGGCAAATGTAT
TAATGGTGAATACATGAGTGATCGTGGCCTTGCATTAAATGACTGCCATGCAGAAATAATATCTCGGAGATCCTTGC
TCAGATTTCTTTATACACAACTTGAGCTTTACTTAAATAACAAAGATGATCAAAAAAGATCCATCTTTCAGAAATCA
GAGCGAGGGGGGTTTAGGCTGAAGGAGAATGTCCAGTTTCATCTGTACATCAGCACCTCTCCCTGTGGAGATGCCAG

AATCTTCTCACCACATGAGCCAATCCTGGAAGAACCAGCAGATAGACACCCAAATCGTAAAGCAAGAGGACAGCTAC
GGACCAAAATAGAGTCTGGTCAGGGGACGATTCCAGTGCGCTCCAATGCGAGCATCCAAACGTGGGACGGGGTGCTG
CAAGGGGAGCGGCTGCTCACCATGTCCTGCAGTGACAAGATTGCACGCTGGAACGTGGTGGGCATCCAGGGATCACT
GCTCAGCATTTTCGTGGAGCCCATTTACTTCTCGAGCATCATCCTGGGCAGCCTTTACCACGGGGACCACCTTTCCA
GGGCCATGTACCAGCGGATCTCCAACATAGAGGACCTGCCACCTCTCTACACCCTCAACAAGCCTTTGCTCAGTGGC
ATCAGCAATGCAGAAGCACGGCAGCCAGGGAAGGCCCCCAACTTCAGTGTCAACTGGACGGTAGGCGACTCCGCTAT
TGAGGTCATCAACGCCACGACTGGGAAGGATGAGCTGGGCCGCGCGTCCCGCCTGTGTAAGCACGCGTTGTACTGTC
GCTGGATGCGTGTGCACGGCAAGGTTCCCTCCCACTTACTACGCTCCAAGATTACCAAGCCCAACGTGTACCATGAG
TCCAAGCTGGCGGCAAAGGAGTACCAGGCCGCCAAGGCGCGTCTGTTCACAGCCTTCATCAAGGCGGGGCTGGGGGC
CTGGGTGGAGAAGCCCACCGAGCAGGACCAGTTCTCACTCACGTACCCATACGACGTACCAGATTACGCTTAA
(SEQ ID NO: 41)

To serve as the target for the putative RNA base-editor, wild-type mCherry coding sequence was mutated to create a premature stop codon TAG (See bold double underlined sequence in SEQ ID NO: 42), such that no functional mCherry protein would be produced without correcting A to G by the RNA base editor. See FIGs. 12 and 14. gRNA was then designed to effect the desired A-to-G editing (FIGs. 12 and 14), and the CX530 plasmid encoding the dCas13e.1-ADAR2DD base editor, the CX537/Cx538 plasmid encoding the sgRNA, and the CX337 plasmid encoding the mutated mCherry gene, were triple transfected into HEK293T cells using standard protocol. Transfected HEK293T cells were incubated for 24 hours at 37°C under 5% CO_2, before the cells were subject to flow cytometry to isolate cells having corrected mCherry mRNA and expressing mCherry protein. See illustrative drawing FIG. 12. The results of flow cytometry analysis were shown in FIG. 13.

It is apparent that both gRNA-1 (SEQ ID NO: 43) and gRNA-2 (SEQ ID NO: 44) successfully corrected the TAG premature stop codon to generate functional mCherry proteins.

ATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTC
CGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGA
AGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTAC
GTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAG**TAG**GAGCGCGTGATGAA
CTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGC
TGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGG
ATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGC
TGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACA
TCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATG
GACGAGCTGTACAAGTAA (SEQ ID NO: 42)

caagtagtcggggatgtcggcggggtgcttcacCtaggccttggagccgtGCTGGAGCAGCCCCCGATTTGTGGGGT
GATTACAGC (SEQ ID NO: 43)

cggggatgtcggcggggtgcttcacCtaggccttggagccgtacatgaacGCTGGAGCAGCCCCCGATTTGTGGGGT
GATTACAGC (SEQ ID NO: 44)

Example 6 Single-Base RNA Editing using Shortened dCas13e.1-ADAR2DD Fusion

In order to determine the minimum size of the dCas13e.1 that can be used in RNA single

base editing, a series of five constructs expressing progressively larger C-terminal deletions of dCas13e.1 were generated, each with 30 fewer residues from the C-terminus (*i.e.*, 30-, 60- 90-, 120, and 150-residue deletions). The resulting constructs were used to create coding sequences for dCas13e.1 fused with the high fidelity adar2 (ADAR2DD) at the respective C-terminus. These constructs were cloned into Vysz15 ("V15") to Vysz-19 ("V19") plasmids (FIG. 15) for use in experiments similar to that in Example 4. In all these constructs, the fusion proteins were expressed from the CMV promoter (pCMV) and enhancer (eCMV), and was immediately downstream of an intron that further enhances protein expression. Two Nuclear Localization Sequences (NLSs) were positioned at the N- and C-terminus of the dCas13e.1 portion of the fusion, and the ADAR2 domain (such as ADAR2DD) was fused to the C-terminal NLS through a NLS linker, and was tagged at the C-terminus by an HA-tag. An EGFP coding sequence under the independent control of the EFS promoter (pEFS) was present downstream of the polyA addition sequence for all plasmids.

Interestingly, it was found that progressive C-terminal deletion steadily increased RNA-base editing activity in the fusion editor, such that the editor with 150 C-terminal residue deletion (in V19) exhibited the highest base editing activity. See FIG. 16. However, 180-residue deletion from the C-terminus appeared to have abolished the base editing activity, suggesting that the maximum / optimal deletion from the C-terminal end of Cas13e.1 is likely between 150-180 residues.

Based on this finding, a series of N-terminal deletion mutants were generated for the dCas13e.1 having 150 C-terminal residue deletion. Seven such N-terminal deletion mutants were generated, with 30-, 60-, 90-, 120-, 150-, 180-, and 210-residue deletions, respectively. See FIG. 17. The results in FIG. 18 showed that the best RNA editing activity was observed in the mutant with 180 N-terminal residue deletion and 150 C-terminal residue deletion, *i.e.*, a total of 330-residue deletion from the 775-residue Cas13e.1 protein, to generate the 445-residue optimal dCas13e.1 for generating the ADAR2DD fusion.

Example 7 Mammalian Endogenous mRNA Knock-down Efficiency Comparison Using Different Cas13 Proteins

This experiment demonstrated that Cas13e and Cas13f proteins, especially Cas13f.1,

were highly efficient in knocking down mammalian endogenous target mRNA, better than the previously identified Cas13 proteins.

Specifically, five plasmids were constructed, each expressing one of the Cas13 proteins, namely Cas13e.1 (SEQ ID NO: 22), Cas13f.1 (SEQ ID NO: 23), LwaCas13a (SEQ ID NO: 45), PspCas13b (SEQ ID NO: 46), and RxCas13d (SEQ ID NO: 47). Each plasmid also encoded the mCherry reporter gene, as well as sgRNA / crRNA coding sequences for the respective Cas13 proteins flanked by two native DR sequences. These sgRNA's were designed to have spacer sequences targeting the ANXA4 mRNA. See SEQ ID NOs: 48-50. As negative control, 5 additional plasmids were constructed, each encoding a non-targeting sgRNA / crRNA instead of the ANXA4-targeting sgRNA/crRNA ("the control NT constructs"). See FIG. 19.

ATGCCCAAGAAGAAGCGGAAGGTGGGATCCATGAAAGTGACCAAGGTCGATGGCATCAGCCACAAGAAGTACATCGA
AGAGGGCAAGCTCGTGAAGTCCACCAGCGAGGAAAACCGGACCAGCGAGAGACTGAGCGAGCTGCTGAGCATCCGGC
TGGACATCTACATCAAGAACCCCGACAACGCCTCCGAGGAAGAGAACCGGATCAGAAGAGAGAACCTGAAGAAGTTC
TTTAGCAACAAGGTGCTGCACCTGAAGGACAGCGTGCTGTATCTGAAGAACCGGAAAGAAAAGAACGCCGTGCAGGA
CAAGAACTATAGCGAAGAGGACATCAGCGAGTACGACCTGAAAAACAAGAACAGCTTCTCCGTGCTGAAGAAGATCC
TGCTGAACGAGGACGTGAACTCTGAGGAACTGGAAATCTTTCGGAAGGACGTGGAAGCCAAGCTGAACAAGATCAAC
AGCCTGAAGTACAGCTTCGAAGAGAACAAGGCCAACTACCAGAAGATCAACGAGAACAACGTGGAAAAAGTGGGCGG
CAAGAGCAAGCGGAACATCATCTACGACTACTACAGAGAGAGCGCCAAGCGCAACGACTACATCAACAACGTGCAGG
AAGCCTTCGACAAGCTGTATAAGAAAGAGGATATCGAGAAACTGTTTTTCCTGATCGAGAACAGCAAGAAGCACGAG
AAGTACAAGATCCGCGAGTACTATCACAAGATCATCGGCCGGAAGAACGACAAAGAGAACTTCGCCAAGATTATCTA
CGAAGAGATCCAGAACGTGAACAACATCAAAGAGCTGATTGAGAAGATCCCCGACATGTCTGAGCTGAAGAAAAGCC
AGGTGTTCTACAAGTACTACCTGGACAAAGAGGAACTGAACGACAAGAATATTAAGTACGCCTTCTGCCACTTCGTG
GAAATCGAGATGTCCCAGCTGCTGAAAAACTACGTGTACAAGCGGCTGAGCAACATCAGCAACGATAAGATCAAGCG
GATCTTCGAGTACCAGAATCTGAAAAAGCTGATCGAAAACAAACTGCTGAACAAGCTGGACACCTACGTGCGGAACT
GCGGCAAGTACAACTACTATCTGCAAGTGGGCGAGATCGCCACCTCCGACTTTATCGCCCGGAACCGGCAGAACGAG
GCCTTCCTGAGAAACATCATCGGCGTGTCCAGCGTGGCCTACTTCAGCCTGAGGAACATCCTGGAAACCGAGAACGA
GAACGATATCACCGGCCGGATGCGGGGCAAGACCGTGAAGAACAACAAGGGCGAAGAGAAATACGTGTCCGGCGAGG
TGGACAAGATCTACAATGAGAACAAGCAGAACGAAGTGAAAGAAAATCTGAAGATGTTCTACAGCTACGACTTCAAC
ATGGACAACAAGAACGAGATCGAGGACTTCTTCGCCAACATCGACGAGGCCATCAGCAGCATCAGACACGGCATCGT
GCACTTCAACCTGGAACTGGAAGGCAAGGACATCTTCGCCTTCAAGAATATCGCCCCCAGCGAGATCTCCAAGAAGA
TGTTTCAGAACGAAATCAACGAAAAGAAGCTGAAGCTGAAAATCTTCAAGCAGCTGAACAGCGCCAACGTGTTCAAC
TACTACGAGAAGGATGTGATCATCAAGTACCTGAAGAATACCAAGTTCAACTTCGTGAACAAAAACATCCCCTTCGT
GCCCAGCTTCACCAAGCTGTACAACAAGATTGAGGACCTGCGGAATACCCTGAAGTTTTTTTGGAGCGTGCCCAAGG
ACAAAGAAGAGAAGGACGCCCAGATCTACCTGCTGAAGAATATCTACTACGGCGAGTTCCTGAACAAGTTCGTGAAA
AACTCCAAGGTGTTCTTTAAGATCACCAATGAAGTGATCAAGATTAACAAGCAGCGGAACCAGAAAACCGGCCACTA
CAAGTATCAGAAGTTCGAGAACATCGAGAAAACCGTGCCCGTGGAATACCTGGCCATCATCCAGAGCAGAGAGATGA
TCAACAACCAGGACAAAGAGGAAAAGAATACCTACATCGACTTTATTCAGCAGATTTTCCTGAAGGGCTTCATCGAC
TACCTGAACAAGAACAATCTGAAGTATATCGAGAGCAACAACAACAATGACAACAACGACATCTTCTCCAAGATCAA
GATCAAAAAGGATAACAAAGAGAAGTACGACAAGATCCTGAAGAACTATGAGAAGCACAATCGGAACAAAGAAATCC
CTCACGAGATCAATGAGTTCGTGCGCGAGATCAAGCTGGGGAAGATTCTGAAGTACACCGAGAATCTGAACATGTTT
TACCTGATCCTGAAGCTGCTGAACCACAAAGAGCTGACCAACCTGAAGGGCAGCCTGGAAAAGTACCAGTCCGCCAA
CAAAGAAGAAACCTTCAGCGACGAGCTGGAACTGATCAACCTGCTGAACCTGGACAACAACAGAGTGACCGAGGACT
TCGAGCTGGAAGCCAACGAGATCGGCAAGTTCCTGGACTTCAACGAAAACAAAATCAAGGACCGGAAAGAGCTGAAA
AAGTTCGACACCAACAAGATCTATTTCGACGGCGAGAACATCATCAAGCACCGGGCCTTCTACAATATCAAGAAATA
CGGCATGCTGAATCTGCTGGAAAAGATCGCCGATAAGGCCAAGTATAAGATCAGCCTGAAAGAACTGAAAGAGTACA
GCAACAAGAAGAATGAGATTGAAAAGAACTACACCATGCAGCAGAACCTGCACCGGAAGTACGCCAGACCCAAGAAG
GACGAAAAGTTCAACGACGAGGACTACAAAGAGTATGAGAAGGCCATCGGCAACATCCAGAAGTACACCCACCTGAA
GAACAAGGTGGAATTCAATGAGCTGAACCTGCTGCAGGGCCTGCTGCTGAAGATCCTGCACCGGCTCGTGGGCTACA

CCAGCATCTGGGAGCGGGACCTGAGATTCCGGCTGAAGGGCGAGTTTCCCGAGAACCACTACATCGAGGAAATTTTC
AATTTCGACAACTCCAAGAATGTGAAGTACAAAAGCGGCCAGATCGTGGAAAAGTATATCAACTTCTACAAAGAACT
GTACAAGGACAATGTGGAAAAGCGGAGCATCTACTCCGACAAGAAAGTGAAGAAACTGAAGCAGGAAAAAAAGGACC
TGTACATCCGGAACTACATTGCCCACTTCAACTACATCCCCCACGCCGAGATTAGCCTGCTGGAAGTGCTGGAAAAC
CTGCGGAAGCTGCTGTCCTACGACCGGAAGCTGAAGAACGCCATCATGAAGTCCATCGTGGACATTCTGAAAGAATA
CGGCTTCGTGGCCACCTTCAAGATCGGCGCTGACAAGAAGATCGAAATCCAGACCCTGGAATCAGAGAAGATCGTGC
ACCTGAAGAATCTGAAGAAAAAGAAACTGATGACCGACCGGAACAGCGAGGAACTGTGCGAACTCGTGAAAGTCATG
TTCGAGTACAAGGCCCTGGAATGA (SEQ ID NO: 45)

ATGCCCAAGAAGAAGCGGAAGGTGGTCGACAACATCCCCGCTCTGGTGGAAAACCAGAAGAAGTACTTTGGCACCTA
CAGCGTGATGGCCATGCTGAACGCTCAGACCGTGCTGGACCACATCCAGAAGGTGGCCGATATTGAGGGCGAGCAGA
ACGAGAACAACGAGAATCTGTGGTTTCACCCCGTGATGAGCCACCTGTACAACGCCAAGAACGGCTACGACAAGCAG
CCCGAGAAAACCATGTTCATCATCGAGCGGCTGCAGAGCTACTTCCCATTCCTGAAGATCATGGCCGAGAACCAGAG
AGAGTACAGCAACGGCAAGTACAAGCAGAACCGCGTGGAAGTGAACAGCAACGACATCTTCGAGGTGCTGAAGCGCG
CCTTCGGCGTGCTGAAGATGTACAGGGACCTGACCAACCACTACAAGACCTACGAGGAAAAGCTGAACGACGGCTGC
GAGTTCCTGACCAGCACAGAGCAACCTCTGAGCGGCATGATCAACAACTACTACACAGTGGCCCTGCGGAACATGAA
CGAGAGATACGGCTACAAGACAGAGGACCTGGCCTTCATCCAGGACAAGCGGTTCAAGTTCGTGAAGGACGCCTACG
GCAAGAAAAAGTCCCAAGTGAATACCGGATTCTTCCTGAGCCTGCAGGACTACAACGGCGACACACAGAAGAAGCTG
CACCTGAGCGGAGTGGGAATCGCCCTGCTGATCTGCCTGTTCCTGGACAAGCAGTACATCAACATCTTTCTGAGCAG
GCTGCCCATCTTCTCCAGCTACAATGCCCAGAGCGAGGAACGGCGGATCATCATCAGATCCTTCGGCATCAACAGCA
TCAAGCTGCCCAAGGACCGGATCCACAGCGAGAAGTCCAACAAGAGCGTGGCCATGGATATGCTCAACGAAGTGAAG
CGGTGCCCCGACGAGCTGTTCACAACACTGTCTGCCGAGAAGCAGTCCCGGTTCAGAATCATCAGCGACGACCACAA
TGAAGTGCTGATGAAGCGGAGCAGCGACAGATTCGTGCCTCTGCTGCTGCAGTATATCGATTACGGCAAGCTGTTCG
ACCACATCAGGTTCCACGTGAACATGGGCAAGCTGAGATACCTGCTGAAGGCCGACAAGACCTGCATCGACGGCCAG
ACCAGAGTCAGAGTGATCGAGCAGCCCCTGAACGGCTTCGGCAGACTGGAAGAGGCCGAGACAATGCGGAAGCAAGA
GAACGGCACCTTCGGCAACAGCGGCATCCGGATCAGAGACTTCGAGAACATGAAGCGGGACGACGCCAATCCTGCCA
ACTATCCCTACATCGTGGACACCTACACACACTACATCCTGGAAAACAACAAGGTCGAGATGTTTATCAACGACAAA
GAGGACAGCGCCCCACTGCTGCCCGTGATCGAGGATGATAGATACGTGGTCAAGACAATCCCCAGCTGCCGGATGAG
CACCCTGGAAATTCCAGCCATGGCCTTCCACATGTTTCTGTTCGGCAGCAAGAAAACCGAGAAGCTGATCGTGGACG
TGCACAACCGGTACAAGAGACTGTTCCAGGCCATGCAGAAAGAAGAAGTGACCGCCGAGAATATCGCCAGCTTCGGA
ATCGCCGAGAGCGACCTGCCTCAGAAGATCCTGGATCTGATCAGCGGCAATGCCCACGGCAAGGATGTGGACGCCTT
CATCAGACTGACCGTGGACGACATGCTGACCGACACCGAGCGGAGAATCAAGAGATTCAAGGACGACCGGAAGTCCA
TTCGGAGCGCCGACAACAAGATGGGAAAGAGAGGCTTCAAGCAGATCTCCACAGGCAAGCTGGCCGACTTCCTGGCC
AAGGACATCGTGCTGTTTCAGCCCAGCGTGAACGATGGCGAGAACAAGATCACCGGCCTGAACTACCGGATCATGCA
GAGCGCCATTGCCGTGTACGATAGCGGCGACGATTACGAGGCCAAGCAGCAGTTCAAGCTGATGTTCGAGAAGGCCC
GGCTGATCGGCAAGGGCACAACAGAGCCTCATCCATTTCTGTACAAGGTGTTCGCCCGCAGCATCCCCGCCAATGCC
GTCGAGTTCTACGAGCGCTACCTGATCGAGCGGAAGTTCTACCTGACCGGCCTGTCCAACGAGATCAAGAAAGGCAA
CAGAGTGGATGTGCCCTTCATCCGGCGGGACCAGAACAAGTGGAAAACACCCGCCATGAAAACCCTGGGCAGAATCT
ACAGCGAGGATCTGCCCGTGGAACTGCCCAGACAGATGTTCGACAATGAGATCAAGTCCCACCTGAAGTCCCTGCCA
CAGATGGAAGGCATCGACTTCAACAATGCCAACGTGACCTATCTGATCGCCGAGTACATGAAGAGAGTGCTGGACGA
CGACTTCCAGACCTTCTACCAGTGGAACCGCAACTACCGGTACATGGACATGCTTAAGGGCGAGTACGACAGAAAGG
GCTCCCTGCAGCACTGCTTCACCAGCGTGGAAGAGAGAGAAGGCCTCTGGAAAGAGCGGGCCTCCAGAACAGAGCGG
TACAGAAAGCAGGCCAGCAACAAGATCCGCAGCAACCGGCAGATGAGAAACGCCAGCAGCGAAGAGATCGAGACAAT
CCTGGATAAGCGGCTGAGCAACAGCCGGAACGAGTACCAGAAAAGCGAGAAAGTGATCCGGCGCTACAGAGTGCAGG
ATGCCCTGCTGTTTCTGCTGGCCAAAAAGACCCTGACCGAACTGGCCGATTTCGACGGCGAGAGGTTCAAACTGAAA
GAAATCATGCCCGACGCCGAGAAGGGAATCCTGAGCGAGATCATGCCCATGAGCTTCACCTTCGAGAAAGGCGGCAA
GAAGTACACCATCACCAGCGAGGGCATGAAGCTGAAGAACTACGGCGACTTCTTTGTGCTGGCTAGCGACAAGAGGA
TCGGCAACCTGCTGGAACTCGTGGGCAGCGACATCGTGTCCAAAGAGGATATCATGGAAGAGTTCAACAAATACGAC
CAGTGCAGGCCCGAGATCAGCTCCATCGTGTTCAACCTGGAAAAGTGGGCCTTCGACACATACCCCGAGCTGTCTGC
CAGAGTGGACCGGGAAGAGAAGGTGGACTTCAAGAGCATCCTGAAAATCCTGCTGAACAACAAGAACATCAACAAAG
AGCAGAGCGACATCCTGCGGAAGATCCGGAACGCCTTCGATCACAACAATTACCCCGACAAAGGCGTGGTGGAAATC
AAGGCCCTGCCTGAGATCGCCATGAGCATCAAGAAGGCCTTTGGGGAGTACGCCATCATGAAGGGATCCCTTCAATG
A (SEQ ID NO: 46)

ATGCCTAAAAAGAAAAGAAAGGTGGGTTCTGGTATCGAGAAGAAGAAGAGCTTCGCCAAGGGCATGGGAGTGAAGAG
CACCCTGGTGTCCGGCTCTAAGGTGTACATGACCACATTTGCTGAGGGAAGCGACGCCAGGCTGGAGAAGATCGTGG
AGGGCGATAGCATCAGATCCGTGAACGAGGGAGAGGCTTTCAGCGCCGAGATGGCTGACAAGAACGCTGGCTACAAG
ATCGGAAACGCCAAGTTTTCCCACCCAAAGGGCTACGCCGTGGTGGCTAACAACCCACTGTACACCGGACCAGTGCA

GCAGGACATGCTGGGACTGAAGGAGACACTGGAGAAGAGGTACTTCGGCGAGTCCGCCGACGGAAACGATAACATCT
GCATCCAGGTCATCCACAACATCCTGGATATCGAGAAGATCCTGGCTGAGTACATCACAAACGCCGCTTACGCCGTG
AACAACATCTCCGGCCTGGACAAGGATATCATCGGCTTCGGAAAGTTTTCTACCGTGTACACATACGACGAGTTCAA
GGATCCAGAGCACCACCGGGCCGCTTTTAACAACAACGACAAGCTGATCAACGCCATCAAGGCTCAGTACGACGAGT
TCGATAACTTTCTGGATAACCCCAGGCTGGGCTACTTCGGACAGGCTTTCTTTTCTAAGGAGGGCAGAAACTACATC
ATCAACTACGGAAACGAGTGTTACGACATCCTGGCCCTGCTGAGCGGACTGAGGCACTGGGTGGTGCACAACAACGA
GGAGGAGTCTCGGATCAGCCGCACCTGGCTGTACAACCTGGACAAGAACCTGGATAACGAGTACATCTCCACACTGA
ACTACCTGTACGACAGGATCACCAACGAGCTGACAAACAGCTTCTCCAAGAACTCTGCCGCTAACGTGAACTACATC
GCTGAGACCCTGGGCATCAACCCAGCTGAGTTCGCTGAGCAGTACTTCAGATTTTCCATCATGAAGGAGCAGAAGAA
CCTGGGCTTCAACATCACAAAGCTGAGAGAAGTGATGCTGGACAGAAAGGATATGTCCGAGATCAGGAAGAACCACA
AGGTGTTCGATTCTATCAGAACCAAGGTGTACACAATGATGGACTTTGTGATCTACAGGTACTACATCGAGGAGGAT
GCCAAGGTGGCCGCTGCCAACAAGAGCCTGCCCGACAACGAGAAGTCTCTGAGCGAGAAGGATATCTTCGTGATCAA
CCTGAGAGGCTCCTTTAACGACGATCAGAAGGACGCTCTGTACTACGATGAGGCCAACAGGATCTGGAGAAAGCTGG
AGAACATCATGCACAACATCAAGGAGTTCCGGGGAAACAAGACCCGCGAGTACAAGAAGAAGGACGCTCCAAGGCTG
CCTAGGATCCTGCCTGCTGGAAGGGACGTGAGCGCCTTCAGCAAGCTGATGTACGCCCTGACAATGTTTCTGGACGG
AAAGGAGATCAACGATCTGCTGACCACACTGATCAACAAGTTCGACAACATCCAGTCTTTTCTGAAAGTGATGCCTC
TGATCGGCGTGAACGCTAAGTTCGTGGAGGAGTACGCCTTCTTTAAGGACAGCGCCAAGATCGCTGATGAGCTGCGG
CTGATCAAGTCCTTTGCCAGGATGGGAGAGCCAATCGCTGACGCTAGGAGAGCTATGTACATCGATGCCATCCGGAT
CCTGGGAACCAACCTGTCTTACGACGAGCTGAAGGCTCTGGCCGACACCTTCAGCCTGGATGAGAACGGCAACAAGC
TGAAGAAGGGCAAGCACGGAATGCGCAACTTCATCATCAACAACGTGATCAGCAACAAGCGGTTTCACTACCTGATC
AGATACGGCGACCCAGCTCACCTGCACGAGATCGCTAAGAACGAGGCCGTGGTGAAGTTCGTGCTGGGACGGATCGC
CGATATCCAGAAGAAGCAGGGCCAGAACGGAAAGAACCAGATCGACCGCTACTACGAGACCTGCATCGGCAAGGATA
AGGGAAAGTCCGTGTCTGAGAAGGTGGACGCTCTGACCAAGATCATCACAGGCATGAACTACGACCAGTTCGATAAG
AAGAGATCTGTGATCGAGGACACCGGAAGGGAGAACGCCGAGAGAGAGAAGTTTAAGAAGATCATCAGCCTGTACCT
GACAGTGATCTACCACATCCTGAAGAACATCGTGAACATCAACGCTAGATACGTGATCGGCTTCCACTGCGTGGAGC
GCGATGCCCAGCTGTACAAGGAGAAGGGATACGACATCAACCTGAAGAAGCTGGAGGAGAAGGGCTTTAGCTCCGTG
ACCAAGCTGTGCGCTGGAATCGACGAGACAGCCCCCGACAAGAGGAAGGATGTGGAGAAGGAGATGGCCGAGAGAGC
TAAGGAGAGCATCGACTCCCTGGAGTCTGCTAACCCTAAGCTGTACGCCAACTACATCAAGTACTCCGATGAGAAGA
AGGCCGAGGAGTTCACCAGGCAGATCAACAGAGAGAAGGCCAAGACCGCTCTGAACGCCTACCTGAGGAACACAAAG
TGGAACGTGATCATCCGGGAGGACCTGCTGCGCATCGATAACAAGACCTGTACACTGTTCCGGAACAAGGCTGTGCA
CCTGGAGGTGGCTCGCTACGTGCACGCCTACATCAACGACATCGCCGAGGTGAACTCCTACTTTCAGCTGTACCACT
ACATCATGCAGAGGATCATCATGAACGAGAGATACGAGAAGTCTAGCGGCAAGGTGTCTGAGTACTTCGACGCCGTG
AACGATGAGAAGAAGTACAACGATAGACTGCTGAAGCTGCTGTGCGTGCCTTTCGGATACTGTATCCCACGGTTTAA
GAACCTGAGCATCGAGGCCCTGTTCGACCGCAACGAGGCTGCCAAGTTTGATAAGGAGAAGAAGAAGGTGAGCGGCA
ACTCCTGA (SEQ ID NO: 47)

ATGGCCCTTCGCAGCTCTTGCACGTCATAC (SEQ ID NO: 48)

TTAGGCAGCCCTCATCAGTGCCGGCTCCCT (SEQ ID NO: 49)

GGCCAGGATCTCAATTAGGCAGCCCTCATC (SEQ ID NO: 50)

The five Cas13 / sgRNA-encoding plasmids were transfected into HEK293 cells as in Example 4. After culturing for 24 hours, cells expressing mCherry were isolated through flow cytometry, and expression of ANXA4 mRNA was determined using RT-PCR to assess knock-down efficiency as compared to control cells transfected by Cas13 / NT-encoding plasmids.

FIG. 20 showed that Cas13b only had marginal ANXA4 mRNA knock-down, while Cas13e.1, Cas13f.1, and Cas13d each had over 80% knock down of the target ANXA4 mRNA. Among them, Cas13e.1 appeared to have the most robust knock-down efficiency.

CLAIMS

1. A Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-Cas complex, comprising:

   (1) an RNA guide sequence comprising a spacer sequence capable of hybridizing to a target RNA, and a direct repeat (DR) sequence 3' to the spacer sequence; and,

   (2) a CRISPR-associated protein (Cas) having an amino acid sequence of any one of SEQ ID NOs: 1-7, or a derivative or functional fragment of said Cas;

   wherein the Cas, the derivative, and the functional fragment of said Cas, are capable of (i) binding to the RNA guide sequence and (ii) targeting the target RNA,

   with the proviso that the spacer sequence is not 100% complementary to a naturally-occurring bacterialphage nucleic acid when the complex comprises the Cas of any one of SEQ ID NOs: 1-7.

2. The CRISPR-Cas complex of claim 1, wherein the DR sequence has substantially the same secondary structure as the secondary structure of any one of SEQ ID NOs: 8-14.

3. The CRISPR-Cas complex of claim 1, wherein the DR sequence is encoded by any one of SEQ ID NOs: 8-14.

4. The CRISPR-Cas complex of claim 1, 2, or 3, wherein the target RNA is encoded by a eukaryotic DNA.

5. The CRISPR-Cas complex of claim 4, wherein the eukaryotic DNA is a non-human mammalian DNA, a non-human primate DNA, a human DNA, a plant DNA, an insect DNA, a bird DNA, a reptile DNA, a rodent DNA, a fish DNA, a worm / nematode DNA, a yeast DNA.

6. The CRISPR-Cas complex of any one of claims 1-5, wherein the target RNA is an mRNA.

7. The CRISPR-Cas complex of any one of claims 1-6, wherein the spacer sequence is between 15-60 nucleotides, between 25-50 nucleotides, or about 30 nucleotides.

8. The CRISPR-Cas complex of any one of claims 1-7, wherein the spacer sequence is 90-100% complementary to the target RNA.

9. The CRISPR-Cas complex of any one of claims 1-8, wherein the derivative comprises conserved amino acid substitutions of one or more residues of any one of SEQ ID NOs: 1-7.

10. The CRISPR-Cas complex of claim 9, wherein the derivative comprises only conserved amino acid substitutions.

11. The CRISPR-Cas complex of any one of claims 1-10, wherein the derivative has identical sequence to wild-type Cas of any one of SEQ ID NOs: 1-7 in the HEPN domain or the RXXXXH motif.

12. The CRISPR-Cas complex of any one of claims 1-9, wherein the derivative is capable of binding to the RNA guide sequence hybridized to the target RNA, but has no RNase catalytic activity due to a mutation in the RNase catalytic site of the Cas.

13. The CRISPR-Cas complex of claim 12, wherein the derivative has an N-terminal deletion of no more than 210 residues, and/or a C-terminal deletion of no more than 180 residues.

14. The CRISPR-Cas complex of claim 13, wherein the derivative has an N-terminal deletion of about 180 residues, and/or a C-terminal deletion of about 150 residues.

15. The CRISPR-Cas complex of any one of claims 12-14, wherein the derivative further comprises an RNA base-editing domain.

16. The CRISPR-Cas complex of claim 15, wherein the RNA base-editing domain is an adenosine deaminase, such as a double-stranded RNA-specific adenosine deaminase (*e.g.*, ADAR1 or ADAR2); apolipoprotein B mRNA editing enzyme; catalytic polypeptide-like (APOBEC); or activation-induced cytidine deaminase (AID).

17. The CRISPR-Cas complex of claim 16, wherein the ADAR2 has E488Q/T375G double mutation or is ADAR2DD.

18. The CRISPR-Cas complex of any one of claims 15-17, wherein the base-editing domain is further fused to an RNA-binding domain, such as MS2.

19. The CRISPR-Cas complex of any one of claims 12-14, wherein the derivative further comprises an RNA methyltransferase, a RNA demethylase, an RNA splicing modifier, a localization factor, or a translation modification factor.

20. The CRISPR-Cas complex of any one of claims 1-19, wherein the Cas, the derivative, or the functional fragment comprises a nuclear localization signal (NLS) sequence or a nuclear export signal (NES).

21. The CRISPR-Cas complex of any one of claims 1-20, wherein targeting of the target RNA results in a modification of the target RNA.

22. The CRISPR-Cas complex of claim 21, wherein the modification of the target RNA is a cleavage of the target RNA.

23. The CRISPR-Cas complex of claim 21, wherein the modification of the target RNA is deamination of an adenosine (A) to an inosine (I).

24. The CRISPR-Cas complex of any one of claims 1-23, further comprising a target RNA comprising a sequence capable of hybridizing to the spacer sequence.

25. A fusion protein, comprising (1) the Cas, the derivative thereof, or the functional fragment thereof, of any one of claims 1-24, and (2) a heterologous functional domain.

26. The fusion protein of claim 25, wherein the heterologous functional domain comprises: a nuclear localization signal (NLS), a reporter protein or a detection label (*e.g.*, GST, HRP, CAT, GFP, HcRed, DsRed, CFP, YFP, BFP), a localization signal, a protein targeting moiety, a DNA binding domain (*e.g.*, MBP, Lex A DBD, Gal4 DBD), an epitope tag (*e.g.*, His, myc, V5, FLAG, HA, VSV-G, Trx, *etc*), a transcription activation domain (*e.g.*, VP64 or VPR), a transcription inhibition domain (*e.g.*, KRAB moiety or SID moiety), a nuclease (*e.g.*, FokI), a deamination domain (*e.g.*, ADAR1, ADAR2, APOBEC, AID, or TAD), a methylase, a demethylase, a transcription release factor, an HDAC, a polypeptide having ssRNA cleavage activity, a polypeptide having dsRNA cleavage activity, a polypeptide having ssDNA cleavage activity, a polypeptide having dsDNA cleavage activity, a DNA or RNA ligase, or any combination thereof.

27. The fusion protein of claim 25 or 26, wherein the heterologous functional domain is fused N-terminally, C-terminally, or internally in the fusion protein.

28. A conjugate, comprising (1) the Cas, the derivative thereof, or the functional fragment thereof, of any one of claims 1-24, conjugated to (2) a heterologous functional moiety.

29. The conjugate of claim 28, wherein the heterologous functional moiety comprises: a nuclear localization signal (NLS), a reporter protein or a detection label (*e.g.*, GST, HRP, CAT, GFP, HcRed, DsRed, CFP, YFP, BFP), a localization signal, a protein targeting moiety, a DNA binding domain (*e.g.*, MBP, Lex A DBD, Gal4 DBD), an epitope tag (*e.g.*, His, myc, V5, FLAG, HA, VSV-G, Trx, *etc*), a transcription activation domain (*e.g.*, VP64 or VPR), a transcription inhibition domain (*e.g.*, KRAB moiety or SID moiety), a nuclease (*e.g.*, FokI), a deamination domain (*e.g.*, ADAR1, ADAR2, APOBEC, AID, or TAD), a methylase, a demethylase, a transcription release factor, an HDAC, a polypeptide having ssRNA cleavage activity, a polypeptide having dsRNA cleavage activity, a polypeptide having ssDNA cleavage activity, a polypeptide having dsDNA cleavage activity, a DNA or RNA ligase, or any combination thereof.

30. The conjugate of claim 28 or 29, wherein the heterologous functional moiety is conjugated N-terminally, C-terminally, or internally with respect to the Cas, the derivative thereof, or the functional fragment thereof.

31. A polynucleotide encoding any one of SEQ ID NOs: 1-7, or a derivative thereof, or a functional fragment thereof, or a fusion protein thereof, provided that the polynucleotide is not any one of SEQ ID NOs: 15-21.

32. The polynucleotide of claim 31, which is codon-optimized for expression in a cell.

33. The polynucleotide of claim 32, wherein the cell is a eukaryotic cell.

34. A non-naturally occurring polynucleotide comprising a derivative of any one of SEQ ID NOs: 8-14, wherein said derivative (i) has one or more (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10) nucleotides additions, deletions, or substitutions compared to any one of SEQ ID NOs: 8-14; (ii) has at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 97% sequence identity to any one of SEQ ID NOs: 8-14; (iii) hybridize under stringent conditions with any one of SEQ ID NOs: 8-14 or any of (i) and (ii); or (iv) is a complement of any of (i) - (iii), provided that the derivative is not any one of SEQ ID NOs: 8-14, and that the derivative encodes an RNA (or is an RNA) that has maintained substantially the same secondary structure as any of the RNA encoded by SEQ ID NOs: 8-14.

35. The non-naturally occurring polynucleotide of claim 34, wherein the derivative functions

as a DR sequence for any one of the Cas, the derivative thereof, or the functional fragment thereof, of any one of claims 1-24.

36. A vector comprising the polynucleotide of any one of claims 31-35.

37. The vector of claim 36, wherein the polynucleotide is operably linked to a promoter and optionally an enhancer.

38. The vector of claim 37, wherein the promoter is a constitutive promoter, an inducible promoter, a ubiquitous promoter, or a tissue specific promoter.

39. The vector of any one of claims 36-38, which is a plasmid.

40. The vector of any one of claims 36-38, which is a retroviral vector, a phage vector, an adenoviral vector, a herpes simplex viral (HSV) vector, an AAV vector, or a lentiviral vector.

41. The vector of claim 40, wherein the AAV vector is a recombinant AAV vector of the serotype AAV1, AAV2, AAV4, AAV5, AAV6, AAV7, AAVrh74, AAV8, AAV9, AAV10, AAV 11, AAV 12, or AAV 13.

42. A delivery system comprising (1) a delivery vehicle, and (2) the CRISPR-Cas complex of any one of claims 1-24, the fusion protein of any one of claims 25-27, the conjugate of any one of claims 28-30, the polynucleotide of any one of claims 31-33, or the vector of any one of claims 36-41.

43. The delivery system of claim 42, wherein the delivery vehicle is a nanoparticle, a liposome, an exosome, a microvesicle, or a gene-gun.

44. A cell or a progeny thereof, comprising the CRISPR-Cas complex of any one of claims 1-24, the fusion protein of any one of claims 25-27, the conjugate of any one of claims 28-30, the polynucleotide of any one of claims 31-33, or the vector of any one of claims 36-41.

45. The cell or progeny thereof of claim 44, which is a eukaryotic cell (*e.g.*, a non-human mammalian cell, a human cell, or a plant cell) or a prokaryotic cell (*e.g.*, a bacteria cell).

46. A non-human multicellular eukaryote comprising the cell of claim 44 or 45.

47. The non-human multicellular eukaryote of claim 46, which is an animal (*e.g.*, rodent or primate) model for a human genetic disorder.

48. A method of modifying a target RNA, the method comprising contacting the target RNA with the CRISPR-Cas complex of any one of claims 1-24, wherein the spacer sequence is complementary to at least 15 nucleotides of the target RNA; wherein the Cas, the derivative, or the functional fragment associates with the RNA guide sequence to form the complex; wherein the complex binds to the target RNA; and wherein upon binding of the complex to the target RNA, the Cas, the derivative, or the functional fragment modifies the target RNA.

49. The method of claim 48, wherein the target RNA is modified by cleavage by the Cas.

50. The method of claim 48, wherein the target RNA is modified by deamination by a derivative comprising a Double-stranded RNA-specific adenosine deaminase.

51. The method of any one of claim 48-50, wherein the target RNA is an mRNA, a tRNA, an rRNA, a non-coding RNA, an lncRNA, or a nuclear RNA.

52. The method of any one of claims 48-51, wherein upon binding of the complex to the target RNA, the Cas, the derivative, and the functional fragment does not exhibit substantial (or detectable) collateral RNase activity.

53. The method of any one of claims 48-52, wherein the target RNA is within a cell.

54. The method of claim 53, wherein the cell is a cancer cell.

55. The method of claim 53, wherein the cell is infected with an infectious agent.

56. The method of claim 55, wherein the infectious agent is a virus, a prion, a protozoan, a fungus, or a parasite.

57. The method of any one of claims 53-56, wherein the CRISPR-Cas complex is encoded by a first polynucleotide encoding any one of SEQ ID NOs: 1-7, or a derivative or functional fragment thereof, and a second polynucleotide comprising any one of SEQ ID NOs: 8-14 and a sequence encoding a spacer RNA capable of binding to the target RNA, wherein the first and the second polynucleotides are introduced into the cell.

58. The method of claim 57, wherein the first and the second polynucleotides are introduced into the cell by the same vector.

59. The method of any one of claims 53-58, which cases one or more of: (i) *in vitro* or *in vivo* induction of cellular senescence; (ii) *in vitro* or *in vivo* cell cycle arrest; (iii) *in vitro* or *in vivo* cell growth inhibition and/or cell growth inhibition; (iv) *in vitro* or *in vitro* induction of anergy; (v) *in vitro* or *in vitro* induction of apoptosis; and (vi) *in vitro* or *in vitro* induction of necrosis.

60. A method of treating a condition or disease in a subject in need thereof, the method comprising administering to the subject a composition comprising the CRISPR-Cas complex of any one of claims 1-24 or a polynucleotide encoding the same; wherein the spacer sequence is complementary to at least 15 nucleotides of a target RNA associated with the condition or disease; wherein the Cas, the derivative, or the functional fragment associates with the RNA guide sequence to form the complex; wherein the complex binds to the target RNA; and wherein upon binding of the complex to the target RNA, the Cas, the derivative or the functional fragment cleaves the target RNA, thereby treating the condition or disease in the subject.

61. The method of claim 60, wherein the condition or disease is a cancer or an infectious disease.

62. The method of claim 61, wherein the cancer is Wilms' tumor, Ewing sarcoma, a neuroendocrine tumor, a glioblastoma, a neuroblastoma, a melanoma, skin cancer, breast cancer, colon cancer, rectal cancer, prostate cancer, liver cancer, renal cancer, pancreatic cancer, lung cancer, biliary cancer, cervical cancer, endometrial cancer, esophageal cancer, gastric cancer, head and neck cancer, medullary thyroid carcinoma, ovarian cancer, glioma, lymphoma, leukemia, myeloma, acute lymphoblastic leukemia, acute myelogenous leukemia, chronic lymphocytic leukemia, chronic myelogenous leukemia, Hodgkin's lymphoma, non-Hodgkin's lymphoma, or urinary bladder cancer.

63. The method of any one of claims 60-62, which is an *in vitro* method, an *in vivo* method, or an *ex vivo* method.

64. A cell or a progeny thereof, obtained by the method of any one of claims 48-59, wherein

the cell and the progeny comprises a non-naturally existing modification (*e.g.*, a non-naturally existing modification in a transcribed RNA of the cell / progeny).

65. A method to detect the presence of a target RNA, the method comprising contacting the target RNA with a composition comprising a fusion protein of any one of claims 25-27, or a conjugate of any one of claims 28-30, or a polynucleotide encoding the fusion protein, wherein the fusion protein or the conjugate comprises a detectable label (*e.g.*, one that can be detected by fluorescence, Northern blot, or FISH) and a complexed spacer sequence capable of binding to the target RNA.

66. A eukaryotic cell comprising a Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-Cas complex, said CRISPR-Cas complex comprising:

(1) an RNA guide sequence comprising a spacer sequence capable of hybridizing to a target RNA, and a direct repeat (DR) sequence 3' to the spacer sequence; and,

(2) a CRISPR-associated protein (Cas) having an amino acid sequence of any one of SEQ ID NOs: 1-7, or a derivative or functional fragment of said Cas;

wherein the Cas, the derivative, and the functional fragment of said Cas, are capable of (i) binding to the RNA guide sequence and (ii) targeting the target RNA.