COMPOSITIONS AND METHODS COMPRISING IMPROVEMENTS OF CRISPR GUIDE RNAS USING THE H1 PROMOTER

CROSS-REFERENCE

This application claims the benefit of U.S. Provisional Application No. 62/358,335, filed July 5, 2016, the entirety of which is hereby incorporated by reference.

BACKGROUND

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) together with cas (CRISPR-associated) genes comprise an adaptive immune system that provides acquired resistance against invading foreign nucleic acids in bacteria and archaea (Barrangou et al. (2007) *Science* 315:1709-12). CRISPR consists of arrays of short conserved repeat sequences interspaced by unique variable DNA sequences of similar size called spacers, which often originate from phage or plasmid DNA (Barrangou et al. (2007) *Science* 315:1709-12; Bolotin et al. (2005) *Microbiology* 151 :2551-61; Mojica et al. (2005) *J. Mol. Evol.* 60:174-82). The CRISPR-Cas system functions by acquiring short pieces of foreign DNA (spacers) which are inserted into the CRISPR region and provide immunity against subsequent exposures to phages and plasmids that carry matching sequences (Barrangou et al. (2007) *Science* 315:1709-12; Brouns et al. (2008) *Science* 321:960-64). It is this CRISPR-Cas interference/immunity that enables crRNA-mediated silencing of foreign nucleic acids (Horvath & Barrangou (2010) *Science* 327:167-70; Deveau et al. (2010) *Annu. Rev. Microbiol.* 64:475-93; Marraffini & Sontheimer (2010) *Nat. Rev. Genet.* 11:181-90; Bhaya et al. (2011) *Annu. Rev. Genet.* 45:273-97; Wiedenheft et al. (2012) *Nature* 482:331-338).

Use of CRISPR constructs that rely upon the nuclease activity of the Cas9 protein (Makarova et al. (2011) *Nat. Rev. Microbiol.* 9:467-77) coupled with a synthetic guide RNA (gRNA) has recently revolutionized genomic-engineering, allowing for unprecedented manipulation of DNA sequences. CRISPR/Cas9 constructs are simple and fast to synthesize and can be multiplexed. However, despite the relative ease of their synthesis, CRISPRs have technological restrictions related to their access to targetable genome space, which is a function of both the properties of Cas9 itself and the synthesis of its gRNA.

Cleavage by the CRISPR system requires complementary base pairing of the gRNA to a 20-nucleotide DNA sequence and the requisite protospacer-adjacent motif (PAM), a

short nucleotide motif found 3' to the target site (Jinek et al. (2012) *Science* 337: 816-821). One can, theoretically, target any unique N_{20}-PAM sequence in the genome using CRISPR technology. The DNA binding specificity of the PAM sequence, which varies depending upon the species of origin of the specific Cas9 employed, provides one constraint. Currently, the least restrictive and most commonly used Cas9 protein is from *S. pyogenes*, which recognizes the sequence NGG, and thus, any unique 21-nucleotide sequence in the genome followed by two guanosine nucleotides (N_{20}NGG) can be targeted. Expansion of the available targeting space imposed by the protein component is limited to the discovery and use of novel Cas9 proteins with altered PAM requirements (Cong et al. (2013) *Science* 339: 819-823; Hou et al. (2013) *Proc. Natl. Acad. Sci. U.S.A.* 110(39):15644-9), or pending the generation of novel Cas9 variants via mutagenesis or directed evolution.

The second technological constraint of the CRISPR system arises from gRNA expression initiating at a 5' guanosine nucleotide. Use of the type III class of RNA polymerase III promoters has been particularly amenable for gRNA expression because these short non-coding transcripts have well-defined ends, and all the necessary elements for transcription, with the exclusion of the 1+ nucleotide, are contained in the upstream promoter region. However, since the commonly used U6 promoter requires a guanosine nucleotide to initiate transcription, use of the U6 promoter has further constrained genomic targeting sites to GN_{19}NGG (Mali et al. (2013) *Science* 339:823-826; Ding et al. (2013) *Cell Stem Cell* 12:393-394 Ranganathan, V *et al. Nature communications* 5, 4516 (2014)). Alternative approaches, such as *in vitro* transcription by T7, T3, or SP6 promoters, would also require initiating guanosine nucleotide(s) (Adhya et al. (1981) *Proc. Natl. Acad. Sci. U.S.A.* 78:147-151; Melton et al. (1984) *Nucleic Acids Res.* 12:7035-7056; Pleiss et al. (1998) *RNA* 4:1313-1317).

SUMMARY

The practice of the present invention will typically employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant nucleic acid (e.g., DNA) technology, immunology, and RNA interference (RNAi) which are within the skill of the art. Non-limiting descriptions of certain of these techniques are found in the following publications: Ausubel, F., et al., (eds.), *Current Protocols in Molecular Biology, Current Protocols in Immunology, Current Protocols in Protein Science*, and *Current Protocols in Cell Biology*, all John Wiley & Sons, N.Y., edition as of December 2008; Sambrook, Russell, and

Sambrook, *Molecular Cloning. A Laboratory Manual,* 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2001; Harlow, E. and Lane, D., Antibodies—A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1988; Freshney, R. I., "Culture of Animal Cells, A Manual of Basic Technique", 5th ed., John Wiley & Sons, Hoboken, N.J., 2005. Non-limiting information regarding therapeutic agents and human diseases is found in Goodman and Gilman's The Pharmacological Basis of Therapeutics, 11th Ed., McGraw Hill, 2005, Katzung, B. (ed.) Basic and Clinical Pharmacology, McGraw-Hill/Appleton & Lange 10th ed. (2006) or 11th edition (July 2009). Non-limiting information regarding genes and genetic disorders is found in McKusick, V. A.: Mendelian Inheritance in Man. A Catalog of Human Genes and Genetic Disorders. Baltimore: Johns Hopkins University Press, 1998 (12th edition) or the more recent online database: Online Mendelian Inheritance in Man, OMIM™. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, Md.) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, Md.), as of May 1, 2010, available on the World Wide Web: http://www.ncbi.nlm.nih.gov/omim/ and in Online Mendelian Inheritance in Animals (OMIA), a database of genes, inherited disorders and traits in animal species (other than human and mouse), available on the World Wide Web: http://omia.angis.org.au/contact.shtml. All patents, patent applications, and other publications (e.g., scientific articles, books, websites, and databases) mentioned herein are incorporated by reference in their entirety. In case of a conflict between the specification and any of the incorporated references, the specification (including any amendments thereof, which may be based on an incorporated reference), shall control. Standard art-accepted meanings of terms are used herein unless indicated otherwise. Standard abbreviations for various terms are used herein.

Provided herein are compositions and methods comprising improvements of a CRISPR/Cas9 system (i.e., CRISPR guide RNAs using the H1 promoter). In some embodiments, the improvements comprise modifications to the H1 promoter region. In some embodiments, the compositions comprise enhancing H1 bidirectional pol II expression using 5'UTR modifications. In some embodiments, the compositions comprise modulating bidirectional expression through use of different orthologous sequences of the H1 promoter. In some embodiments, the compositions comprise novel compact bidirectional promoter sequences with both pol II and pol III activity (e.g., 7sk, 5'UTRs, Kozak consensus sequences, or combinations thereof). In some embodiments, the method

comprises an expression screen for bidirectional promoters with both RNA pol II and RNA pol III activities. In some embodiments, the compositions comprise conditional pol II/pol III bidirectional promoter expression (e.g., TetR and TetO sites) which can regulate ribonucleoprotein enzymatic activity or RNA-directed nucleases. In some embodiments, the improvements comprise addition of a donor template sequence for correcting mutations (e.g., homology directed repair (HDR)).

In other embodiments of the present invention, the improvements comprise modifications to a component of the CRISPR/Cas9 system. In some embodiments, the modifications are made to nucleases (e.g., RNA-guided nucleases). In some embodiments, the nuclease (e.g., Cas9) is modified through post-transcriptional cell-cycle regulation (e.g., fusion proteins comprising Geminin (Gem) or Cdt1). In some embodiments, the nuclease (e.g., Cas9) is modified by engineering partial target sites such that the nuclease can bind without DNA cleavage. In some embodiments, the nuclease (e.g., Cas9) is modified by modulating its half-life using N-terminal amino acid identity.

One aspect of the invention relates to a non-naturally occurring nuclease system comprising one or more vectors comprising: a) a promoter operably linked to at least one nucleotide sequence encoding a nuclease system guide RNA (gRNA), wherein the gRNA hybridizes with a target sequence of a DNA molecule or RNA molecule in a cell, and wherein the DNA molecule or RNA encodes one or more gene products expressed in the cell; and b) a regulatory element operable in a cell operably linked to a nucleotide sequence encoding a genome-targeted nuclease,
wherein components (a) and (b) are located on the same or different vectors of the system, wherein the gRNA targets and hybridizes with the target sequence and the nuclease cleaves the DNA molecule or RNA to alter expression of the one or more gene products.

In some embodiments, the system is CRISPR (*e.g.* CRISPR associated (Cas) 9 (CRISPR-Cas9, non-Cas9 CRISPR systems, CRISPR-Cpf-1 system, and the like).

In some embodiments, the system is packaged into a single adeno-associated virus (AAV) particle.

In some embodiments, the adeno-associated packaging virus is selected from adenovirus serotype 2, adenovirus serotype 5, or adenovirus serotype 35.

In some embodiments, the adeno-associated packaging virus is adenovirus serotype 5.

In some embodiments, the system inactivates one or more gene products.

In some embodiments, the nuclease system excises at least one gene mutation.

In some embodiments, the promoter is selected from the group consisting of H1 promoter, 7sk, human RPPH1-PARP2, SRP-RPS29, 7sk1-GSTA4, SNAR-G-1-CGB1, SNAR- CGB2, RMRP-CCDC107, tRNA(Lys)-ALOXE3, RNU6-9-MED16: tRNA (Gly)-DPP9, RNU6-2-THEM259, SNORD13-C8orf41, mouse RPPH1-PARP2, and rat RPPH1-PARP2.

In some embodiments, the promoter is a H1 promoter.

In some embodiments, the H1 promoter comprises a nucleotide sequence having at least 80% identity to the nucleotide sequence set forth in SEQ ID NO: 12 or 32.

In some embodiments, the H1 promoter comprises a nucleotide sequence having the nucleotide sequence set forth in SEQ ID NO: 12.

In some embodiments, the H1 promoter comprises a nucleotide sequence having the nucleotide sequence set forth in SEQ ID NO: 32.

In some embodiments, the promoter is orthologous to the H1 promoter.

In some embodiments, the orthologous H1 promoter is derived from eutherian mammals.

In some embodiments, the orthologous H1 promoter is derived from *ailuropoda melanoleuca, bos taurus, callithrix jacchus, canis familiaris, cavia porcellus, chlorocebus sabaeus, choloepus hoffmanni, dasypus novemcinctus, dipodomys ordii, equus caballus, erinaceus europaeus, felis catus, gorilla gorilla, homo sapiens, ictidomys tridecemlineatus, loxodonta africana, macaca mulatta, mus musculus, mustela putorius furo, myotis lucifugus, nomascus leucogenys, ochotona princeps, oryctolagus cuniculus, otolemur garnettii, ovis aries, pan troglodytes, papio anubis, pongo abelii, procavia capensis, pteropus vampyrus, rattus norvegicus, sus scrofa, tarsius syrichta, tupaia belangeri, tursiops truncatus, vicugna pacos*.

In some embodiments, the orthologous H1 promoter is derived from mouse or rat.

In some embodiments, the orthologous H1 promoter comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NOs: 84-119.

In some embodiments, the orthologous H1 promoter comprises a nucleotide sequences set forth in the group consisting of SEQ ID NOs: 84-119.

In some embodiments, the H1 promoter is bidirectional. The H1 promoter is both a pol II and pol III promoter

In some embodiments, the H1 promoter comprises: a) control elements that provide for transcription in one direction of at least one nucleotide sequence encoding a gRNA; and b) control elements that provide for transcription in the opposite direction of a nucleotide sequence encoding a genome-targeted nuclease.

In some embodiments, the promoter is a 7sk promoter.

In some embodiments, the 7sk promoter is derived from human, rat, or mouse.

In some embodiments, the 7sk promoter is selected from the group consisting of 7sk1, 7sk2, and 7sk3.

In some embodiments, the 7sk promoter is 7sk1.

In some embodiments, the 7sk1 is derived from human.

In some embodiments, the human 7sk1 comprises a nucleotide sequence having at least 80% identity to the nucleotide sequence set forth in SEQ ID NO: 3.

In some embodiments, the human 7sk1 comprises a nucleotide sequence having the nucleotide sequence set forth in SEQ ID NO: 3.

In some embodiments, the 7sk1 is derived from mouse.

In some embodiments, the mouse 7sk1 comprises a nucleotide sequence having at least 80% identity to SEQ ID NO: 6.

In some embodiments, the mouse 7sk1 comprises a nucleotide sequence having the nucleotide sequence set forth in SEQ ID NO: 6.

In some embodiments, the 7sk promoter is 7sk2.

In some embodiments, the 7sk2 is derived from human.

In some embodiments, the human 7sk2 comprises a nucleotide sequence having at least 80% identity to SEQ ID NO: 4.

In some embodiments, the human 7sk2 comprises a nucleotide sequence having the nucleotide sequence set forth in SEQ ID NO: 4.

In some embodiments, the 7sk2 is derived from mouse.

In some embodiments, the mouse 7sk2 comprises a nucleotide sequence having at least 80% identity to SEQ ID NO: 7.

In some embodiments, the mouse 7sk2 comprises a nucleotide sequence having the nucleotide sequence set forth in SEQ ID NO: 7.

In some embodiments, the 7sk promoter is 7sk3.

In some embodiments, the 7sk3 is derived from human.

In some embodiments, the human 7sk3 comprises a nucleotide sequence having at least 80% identity to SEQ ID NO: 5.

In some embodiments, the human 7sk3 comprises a nucleotide sequence having the nucleotide sequence set forth in SEQ ID NO: 5.

In some embodiments, the 7sk3 is derived from mouse.

In some embodiments, the mouse 7sk3 comprises a nucleotide sequence having at least 80% identity to SEQ ID NO: 8.

In some embodiments, the mouse 7sk3 comprises a nucleotide sequence having the nucleotide sequence set forth in SEQ ID NO: 8.

In some embodiments, the promoter has at least one modification.

In some embodiments, the at least one modification of the promoter comprises an element that allows conditional regulation.

In some embodiments, the element is a tet-responsive promoter.

In some embodiments, the tet-response promoter comprises a Tet repressor (TetR) and Tet operator (TetO) engineered into the H1 promoter.

In some embodiments, the TetR comprises an amino acid sequence having at least 80% identity to SEQ ID NO: 14.

In some embodiments, the TetR comprises an amino acid sequence having the amino acid sequence set forth in SEQ ID NO: 14.

In some embodiments, the TetO comprises a nucleotide sequence having at least 80% identity to the nucleotide sequence set forth in SEQ ID NO: 11.

In some embodiments, the TetO comprises a nucleotide sequence having the nucleotide sequence set forth in SEQ ID NO: 11.

In some embodiments, the H1-TetO comprises a nucleotide sequence having at least 80% identity to the nucleotide sequence set forth in SEQ ID NO: 13.

In some embodiments, the H1-TetO comprises a nucleotide sequence having the nucletoide sequence set forth in SEQ ID NO: 13.

In some embodiments, the at least one modification of the promoter comprises a site that allows auto-regulation.

In some embodiments, the auto-regulation site comprises a nucleotide sequence having at least 80% identity to the nucleotide sequence set forth in SEQ ID NO: 33 or 34.

In some embodiments, the auto-regulation site comprises the nucleotide sequence set forth in SEQ ID NO: 33.

In some embodiments, the auto-regulation site comprises the nucleotide sequence set forth in SEQ ID NO: 34.

In some embodiments, the method further comprises at least one terminator sequence.

In some embodiments, the at least one terminator sequence is selected from the group consisting of SV40 or synthetic poly A (SPA) sequences.

In some embodiments, the terminator sequences is a SV40 120 (SEQ ID NO: 123) or 240 (SEQ ID NO: 122) base pair sequence.

In some embodiments, the terminator sequences is a SPA 49 base pair equence.

In some embodiments, the method further comprises 5' untranslated region (5'UTR) sequences.

In some embodiments, the method further comprises a Kozak sequence.

In some embodiments, the Kozak sequence comprises a nucleotide sequence having at least 80% identity to SEQ ID NO: 1 or 2.

In some embodiments, the Kozak sequence comprises the nucleotide sequence set forth in SEQ ID NO: 1.

In some embodiments, the Kozak sequence comprises the nucleotide sequence set forth in SEQ ID NO: 2.

In some embodiments, the method further comprises a RNA sequence that mediates cap-independent initiation of translation.

In some embodiments, the RNA sequence is selected from the group consisting of 6.947 or 6.967.

In some embodiments, the genome-targeted nuclease is Cas9 protein.

In some embodiments, the Cas9 comprises a nucleotide sequence having at least 80% identity to the nucleotide sequence set forth in SEQ ID NO: 61.

In some embodiments, the Cas9 comprises the nucleotide sequence set forth in SEQ ID NO: 61.

In some embodiments, the Cas9 comprises at least one modification.

In some embodiments, the at least one modification in the Cas9 comprises an alteration in the cleaving sequence.

In some embodiments, the alteration in the cleaving sequence selected from the group consisting of T2A, P2A, E2A, and F2A.

In some embodiments, the T2A comprises an amino acid sequence. having at least 80% identity to the nucleotide sequence set forth in SEQ ID NO: 36.

In some embodiments, the T2A comprises an amino acid sequence having the nucleotide sequence set forth in SEQ ID NO: 36.

In some embodiments, the P2A comprises an amino acid sequence having at least 80% identity to the nucleotide sequence set forth in SEQ ID NO: 35.

In some embodiments, the P2A comprises an amino acid sequence having the nucleotide sequence set forth in SEQ ID NO: 35.

In some embodiments, the E2A comprises an amino acid sequence having at least 80% identity to the nucleotide sequence set forth in SEQ ID NO: 37.

In some embodiments, the E2A comprises an amino acid sequence having the nucleotide sequence set forth in SEQ ID NO: 37.

In some embodiments, the F2A comprises an amino acid sequence having at least 80% identity to the nucleotide sequence set forth in SEQ ID NO: 38.

In some embodiments, the F2A comprises an amino acid sequence having the nucleotide sequence set forth in SEQ ID NO: 38.

In some embodiments, the at least one modification in the Cas9 comprises a codon optimized for expression in the cell.

In some embodiments, the Cas9 further comprises a linker sequence operably fused in frame to a cell-cycle dependent protein (Cas9 fusion).

In some embodiments, the cell cycle-dependent protein is selected from the group consisting of APE2, ATR, BRCA1, Chk1, Cdc5, Cdc6, Cdc7, Cdc45, Cdt1, CSA, CSB, Ctf18, DDB1, DDB2, DNA2, DUT, Elg1, EndoV, Esp1, Exonuclease1, FBH1, FEN1, Geminin, Hus1, KNTC2 (NDC80), Ku80, Ligase1, Mad2, MBD4, Mcm3, Mcm4, Mcm5, Mcm6, Mcm7, Mcm8, Mcm10, MGMT, MLH3, Mms4, MPG, MSH2, Mus81, NBS1, NEIL2, NEIL3, NTH1, Orc1, Orc3, PARP1, PCNA, Pif1, PMS1, PMS2, PNK, Pola p180, Pola p70, Pola Spp1 (Prim2a), Polb, Pold p125, Pole Dpb3, Pole Dpb4, Pole Pol2, Poli, Poll, Polm, Psf1, Psf2, Psf3, Rad1, Rad18, Rad23A, Rad23B, Rad51, Rad51D, Rad54, Rad6A, RPA34, RPA70, Scc1, Scc3, Sir2, SIRT1 (Sirtuin), TDG, TDP1, TIMELESS, Tin2, Topoisomerase I, Topoisomerase IIIa, Topoisomerase IIIb, Ubc13, UNG, XAB2, XPC, XPF, XPG, Xrcc2, and XRCC4.

In some embodiments, the cell cycle-dependent protein is Geminin.

In some embodiments, the cell cycle-dependent protein is human Geminin.

In some embodiments, the human Geminin comprises the amino acids from positions 1-110 (hGem(1-110)).

In some embodiments, the hGem(1-110) comprises an amino acid sequence having at least 80% identity to the amimo acid sequence set forth in SEQ ID NO: 19.

In some embodiments, the hGem(1-110) comprises the amino acid sequence set forth in SEQ ID NO: 19.

In some embodiments, the cell cycle-dependent protein is Cdt1.

In some embodiments, the cell cycle-dependent protein is human Cdt1.

In some embodiments, the human Cdt1 comprises amino acids from positions 30-120 (hCdt1(30-120)).

In some embodiments, the hCdt1(30-120) comprises an amino acid sequence having at least 80% identity to the amino acid sequence set forth in SEQ ID NO: 18.

In some embodiments, the hCdt1(30-120) comprises the amino acid sequence set forth in SEQ ID NO: 18.

In some embodiments, the Cas9 fusion comprises an amino acid sequence having at least 80% identity to the amino acid sequence selected from the group consisting of SEQ ID NO: 20-27.

In some embodiments, the Cas9 fusion comprises the amino acid sequence selected from the group consisting of SEQ ID NO: 20-27.

In some embodiments, the Cas9 is operably fused in frame to a ubiquitin protein (Ub-Cas9).

In some embodiments, the Ub-Cas9 at least one N-terminal modification.

In some embodiments, the N-terminal modified Ub-Cas9 comprises an amino acid sequence having at least 80% identity to the amino acid sequence selected from the group consisting of SEQ ID NO: 39-58

In some embodiments, the N-terminal modified Ub-Cas9 comprises the amino acid sequence selected from the group consisting of SEQ ID NO: 39-58.

In some embodiments, the ubiquitin protein comprises a nucleotide sequence having at least 80% identity to the nucleotide sequence set forth in SEQ ID NO: 60.

In some embodiments, the ubiquitin protein comprises the nucleotide sequence set forth in SEQ ID NO: 60.

In some embodiments, the method further comprises a SaCas9 nickase.

In some embodiments, the method further comprisesa donor template sequence.

In some embodiments, the at least one vector comprises a nucleotide sequence having at least 80% identity to the nucleotide sequence set forth in SEQ ID NO: 62 or 67.

In some embodiments, the at least one vector comprises the nucleotide sequence set forth in SEQ ID NO: 62.

In some embodiments, the at least one vector comprises the nucleotide sequence set forth in SEQ ID NO: 67.

In some embodiments, the donor template sequence corrects at least one gene mutation.

In some embodiments, the at least one gene mutation is *rd10* or *rd12*.

In some embodiments, the at least one vector comprises a nucleotide sequence having at least 80% identity to the nucleotide sequence set forth in SEQ ID NO: 63 or 64.

In some embodiments, the at least one vector comprises the nucleotide sequence set forth in SEQ ID NO: 63.

In some embodiments, the at least one vector comprises the nucleotide sequence set forth in SEQ ID NO: 64.

In some embodiments, the at least one vector comprises a nucleotide sequence having at least 80% identity to the nucleotide sequence set forth in SEQ ID NO: 65 or 66.

In some embodiments, the at least one vector comprises the nucleotide sequence set forth in SEQ ID NO: 65.

In some embodiments, the at least one vector comprises the nucleotide sequence set forth in SEQ ID NO: 66.

In some embodiments, the promoter is operably linked to at least one, two, three, four, five, six, seven, eight, nine, or ten gRNA.

In some embodiments, the target sequence comprises the nucleotide sequence $AN_{19}NGG$, $GN_{19}NGG$, $CN_{19}NGG$, or $TN_{19}NGG$.

In some embodiments, the cell is a eukaryotic or or non-eukaryotic cell.

In some embodiments, the eukaryotic cell is a mammalian or human cell.

In some embodiments, the eukaryotic cell is a retinal photoreceptor cell.

In some embodiments, the one or more gene products are rhodopsin.

In some embodiments, the expression of the one or more gene products is decreased.

Another aspect of the invention relates to amethod of altering expression of one or more gene products in a cell, wherein the cell comprises a DNA molecule encoding the one

or more gene products, the method comprising introducing into the cell a non-naturally occurring nuclease system as set forth above.

In some embodiments, the system inactivates one or more gene products.

In some embodiments, the nuclease system excises at least one gene mutation.

In some embodiments, the expression of the one or more gene products is decreased.

In some embodiments, the cell is a eukaryotic or non-eukaryotic cell.

In some embodiments, the eukaryotic cell is a mammalian or human cell.

In some embodiments, the cell is a retinal photoreceptor cell.

In some embodiments, the cell is a retinal ganglion cell.

130. In some embodiments, the eukaryotic cell is a cancerous cell.

In some embodiments, cell proliferation is inhibited or reduced in the cancerous cell.

In some embodiments, apoptosis is enhanced or increased in the cancerous cell.

Another aspect of the invention relates to a method for treating a disease selected from the group consisting of retinal dystrophy, corneal dystrophy, microsatellite expansion diseases, and cancer in a subject in need thereof, the method comprising: (a) providing a non-naturally occurring nuclease system as set forth above; and (b) administering to the subject an effective amount of the system.

In some embodiments, the disease is corneal dystrophy.

In some embodiments, the disease is a retinal dystrophy.

In some embodiments, the retinal dystrophy is selected from the group consisting of Leber's congenital amaurosis (LCA), retinitis pigmentosa (RP), and glaucoma.

In some embodiments, the corneal dystrophy is selected from the group consisting of Epithelial Basement Membrane Dystrophy, Epithelial Recurrent Erosion Dystrophies, Subepithelial Mucinous Corneal Dystrophy, Meesmann Corneal Dystrophy, Lisch Epithelial Corneal Dystrophy, Gelatinous Drop-like Corneal Dystrophy, Reis–Bucklers Corneal Dystrophy, Thiel–Behnke Corneal Dystrophy, Lattice Corneal Dystrophy, Type 1 (Classic), Lattice Corneal Dystrophy, Type 2, Lattice Corneal Dystrophy, Type III, Lattice Corneal Dystrophy, Type IIIA, Lattice Corneal Dystrophy, Type I/IIIA, Lattice Corneal Dystrophy, Type IV, Polymorphic (Corneal) Amyloidosis, Granular Corneal Dystrophy, Type 1, Granular Corneal Dystrophy, Type 2, Macular Corneal Dystrophy, Schnyder Corneal Dystrophy, Congenital Stromal Corneal Dystrophy, Fleck Corneal Dystrophy,

Posterior Amorphous Corneal Dystrophy, Central Cloudy Dystrophy of Francois, Pre-Descemet Corneal Dystrophy, Fuchs Endothelial Corneal Dystrophy, Posterior Polymorphous Corneal Dystrophy, Congenital Hereditary Endothelial Dystrophy, and X-linked Endothelial Corneal Dystrophy.

In some embodiments, the microsatellite expansion diseases is selected from the group consisting of Blepharophimosis, ptosis and epicanthus inversus syndactyly, Cleidocranial dysplasia, Congenital central hypoventilation syndrome, Haddad syndrome DM (Myotonic dystrophy), FRAXA (Fragile X syndrome), FRAXE (Fragile XE mental retardation), FRDA (Friedreich's ataxia), Fuchs' Endothelial Corneal Dystrophy, FXTAS (Fragile X-associated tremor/ataxia syndrome), Hand-foot-genital syndrome, HD (Huntington's disease), Holoprosencephaly, Mental retardation with growth hormone deficiency, Mental retardation, epilepsy, West syndrome, Partington syndrome, Oculopharyngeal muscular dystrophy, SBMA (Spinal and bulbar muscular atrophy), SCA1 (Spinocerebellar ataxia Type 1), SCA12 (Spinocerebellar ataxia Type 12), SCA17 (Spinocerebellar ataxia Type 17), SCA2 (Spinocerebellar ataxia Type 2), SCA3 (Spinocerebellar ataxia Type 3 or Machado-Joseph disease), SCA6 (Spinocerebellar ataxia Type 6), SCA7 (Spinocerebellar ataxia Type 7), SCA8 (Spinocerebellar ataxia Type 8), and Synpolydactyly.

In some embodiments, administering to the subject occurs by implantation, injection, or virally.

In some embodiments, administering to the subject occurs by subretinal injection.

In some embodiments, administering to the subject occurs topically, intravascularly, intradermally, transdermally, parenterally, intravenously, intramuscularly, intranasally, subcutaneously, regionally, percutaneously, intratracheally, intraperitoneally, intraarterially, intravesically, intratumorally, peritumorally, inhalationly, systematically, perfusionly, lavagely, directly via injection, or orally via administration and formulation.

In some embodiments, administering to the subject occurs topically to the surface of the eye.

In some embodiments, administering to the subject occurs on or outside the cornea, sclera, to the intraocular, subconjunctival, sub-tenon, or retrobulbar space, or in or around the eyelids.

In some embodiments, the subject is treated with at least one additional anti-cancer agent.

In some embodiments, the anti-cancer agent is selected from the group consisting of paclitaxel, cisplatin, topotecan, gemcitabine, bleomycin, etoposide, carboplatin, docetaxel, doxorubicin, topotecan, cyclophosphamide, trabectedin, olaparib, tamoxifen, letrozole, and bevacizumab.

In some embodiments, the subject is treated with at least one additional anti-cancer therapy.

In some embodiments, the anti-cancer therapy is radiation therapy, chemotherapy, or surgery.

In some embodiments, the cancer is a solid tumor.

In some embodiments, the cancer is selected from the group consisting of brain cancer, gastrointestinal cancer, oral cancer, breast cancer, ovarian cancer, prostate cancer, pancreatic cancer, lung cancer, liver cancer, throat cancer, stomach cancer, and kidney cancer.

In some embodiments, the cancer is brain cancer.

In some embodiments, the systematic administration is selected from the group consisting of oral, intravenous, intradermal, intraperitoneal, subcutaneous, and intramuscular administration.

In some embodiments, the subject is a mammal.

In some embodiments, the mammal is human.

Certain aspects of the presently disclosed subject matter having been stated hereinabove, which are addressed in whole or in part by the presently disclosed subject matter, other aspects will become evident as the description proceeds when taken in connection with the accompanying Examples and Figures as best described herein below.

BRIEF DESCRIPTION OF THE FIGURES

Having thus described the presently disclosed subject matter in general terms, reference will now be made to the accompanying Figures, which are not necessarily drawn to scale, and wherein:

FIG. 1 comprises six panels, A-F showing the effects of 5'UTR sequences on H1 Pol II expression. **FIG. 1A-1C** show the effects of 5'UTR and translation enhancerson H1 Pol II expression. **FIG. 1D** and **1E** show the effects of 5'UTR Kozak sequences on H1 Pol II expression. FIG. 1F depicts terminator sequences. The functional terminators were roughly equivalent. The shortest effective terminator is the 49bp SPA terminator.

FIG. 2 comprises three panels, A, B, and C showing modulating bidirectional expression through the use of different orthologous sequences.

FIG. 3 comprises two panels, A and B showing H1 alignments of different orghologous sequences.

FIG. 4 comprises two panels, A and B showing a screen for Pol II and Pol III mutants from bidirectional promoters.

FIG. 5 shows further H1 alignments of different orghologous sequences.

FIG. 6 comprises two panels, A and B, showing conditional expression of H1 Pol III expression using TetO on regulating ribonucleoprotein enzymatic activity. **FIG. 6A** shows a schematic of the engineered TetO sites in the H1 promoter. **FIG. 6B** shows that Pol II expression from engineered H1 is not greatly affected by the presence of Tet operator sequences.

FIG. 7 comprises two panels, A and B, showing conditional expression of H1 Pol III expression regulating ribonucleoprotein enzymatic activity. **FIG. 7A** shows that Pol III expression is repressed in the presence of TetR and absence of doxycycline. **FIG. 7B** shows that Pol III expression is derepressed in the presence of TetR and presence of doxycycline.

FIG. 8 comprises two panels, A and B, showing show conditional expression of H1 Pol III expression regulating ribonucleoprotein enzymatic activity. **FIG. 8A** shows the active enzyme complex in the presence of doxycycline. **FIG. 8B** shows the inactive enzyme complex in the absence of doxycycline.

FIG. 9 comprises three panels, A, B, and C, showing a schematic of the cellular response to DNA breaks. **FIG. 9A** shows that the cellular response to DNA breaks is modulated by the cell-cycle. **FIG. 9B** shows that the phase of the cell-cycle largely dictates the choice of DNA pathway. **FIG. 9C** shows thatRNA-directed nucleases can be fused to either hGEM1 or hCDT1 domains, resulting in cell-cycle dependent regulation of these proteins.

FIG. 10 comprises three panels, A, B, and C. **FIG. 10A** shows the cell-cycle regulated construct that is active during S, G2 and M phase (top), and the cell-cycle regulated construct that is active during G1 phase (bottom). **FIG. 10B** shows the oligo sequence used to knock-in an EcoRI site into the Rhodopsin gene, and the location of the CRISPR target and cut site. **FIG. 10C** shows the sequence after successful HDR into the Rhodopsin gene and the incorporation of an EcoRI site into the gene.

FIG. 11 comprises three panels, A, B, and C. **FIG. 11A** shows the quantification of NHEJ using T7 EndoI assay for the different cell-cycle regulated constructs and with different amounts of the donor oligo sequence. **FIG. 11B** shows the quantification of HDR using EcoRI assay for the different cell-cycle regulated constructs and with different amounts of the donor oligo sequence. **FIG. 11C** shows the quantification of the NHEJ and HDR data. The Gem constructs preferentially induce HDR while Cdt constructs preferentially induce NHEJ. Mixture of both (gem/cdt) is similar Cas9 with no fusion, and the overall rate of NHEJ and HDR is approximately equal across all conditions, as is expected.

FIG. 12 comprises two panels, A and B, showing auto-regulation of RNA-guided nucleases using partial target sites. **FIG. 12A** shows Cas9 with perfect complementarity at a target site results in DNA cleavage (white arrows). **FIG. 12B** shows Cas9 with extensive base pairing but without complementarity still binds, however there is no DNA cleavage activity.

FIG. 13 comprises three panels, A, B, and C, showing auto-regulation of RNA-guided nucleases using partial target sites. **FIG. 13A** shows polymerase binding to a 20 base target sequence. **FIG. 13B** shows dCas9 (nuclease-dead version of Cas9), binds to DNA but does not cut. When bound to a promoter region and areas of active transcription, this can inhibit transcription, likely due to steric effects. **FIG. 13C** shows Cas9 with extensive base pairing but without complementarity still binds, however there is no DNA cleavage activity.

FIG. 14 shows auto-regulation of RNA-guided nucleases using partial target sites. Perfect complementarity at desired site results in DNA cleavage (white arrows) (left bottom depiction). Imperfect complementarity at engineered promoter site results in no DNA cleavage and suppression of its own expression (right bottom depiction).

FIG. 15 shows auto-regulation of RNA-guided nucleases using partial target sites:GFP reporter. The bar graphs show that Pol II expression from engineered H1 is not greatly affected by presence of partial target sequences.

FIG. 16 shows auto-regulation of RNA-guided nucleases using partial target sites:Cas9 and gRNA.

FIG. 17 comprises two panels, A and B. **FIG. 17A** shows Methionine (Met) followed by a glycine (Gly), proline (Pro), serine (Ser), threonine (Thr), alanine (Ala),

valine (Val), or cysteine (Cys) are processed by Methionineaminopeptidases, which cleave off the N-terminal methionine. **FIG. 17B** shows that the N-end rule serves as an estimation of a proteins half-life within a cell

FIG. 18 shows that linear poly-ubiquitin is recognized in the cell by deubiquitin enzymes which cleave the individual Ub peptides. This process can be co-opted to generate specific N-terminal residues by fusing Ub to the N-terminus of any protein.

FIG. 19 comprises two panels, A and B, showing that simple changes in the N-terminal amino acids, either through Met-aminopeptidases, deubiquintation, or alternative methods, can accomplish the regulation of RNA-guided nuclease half-life. **FIG. 19A** and **19B** show that the levels of Cas9, an RNA-guided nuclease, can be modulated by the identity of the N-terminal amino acid, by as much as 8-10-fold, a significant range of expression for a nuclease.

FIG. 20 shows how commonly used 2A peptides can also be leveraged to modulate protein levels in the cell.

FIG. 21 comprises three panels, A-C, depicting an illustration of the AAV packaging capacity. Wildtype AAV is ~4.7 kb and recombinant AAV can be stuffed up to 5.2kb. **FIG. 21B** depicts an illustration of AAV virus size using the H1 bidirectional promoter to express SpCas9 and a single gRNA. **FIG. 21C** depicts a construct using the stronger mouse promoter. The first set of experiments were carried out using the human promoter. **FIG. 21B** constructs used the SV40 terminator which was replaced with the smaller SPA terminator

FIG. 22 comprises nine panels, A-I. **FIG. 22A** depicts the AAV delivery problem. The DNA required to express Cas9 and the gRNA, by conventional methods, exceeds 5.2kb: Pol II promoter (~0.5kb), SpCas9 (~4.1kb), Pol II terminator (~0.2kb), U6 promoter (~0.3kb), and the gRNA (~0.1kb). **FIG. 22B** shows the *S. aureus* Cas9 is ~1kb smaller than the *S. pyogenes* Cas9 which allows it to be packaged into AAV. **FIG. 22C** depicts the H1 genomic locus indicating the H1RNA (a pol III transcript) expressed in the forward orientation, and the PARP-2 gene (a pol II transcript) transcribed in the opposite orientation. This ~200bp sequence is an extremely compact bidirectional promoter. **FIG. 22D** depicts shrinking the "instructions" to delliver SpCas9, the most-commonly used and most-widely studied Cas9 protein. Importantly, the SpCas9 protein can target a far greater numbers of genomic sites than the SaCas9 protein. This means that more mutations/diseases can be targeted. **FIG. 22E** shows shrinking the "instructions" to deliver a number of other Cas9 proteins that have been shown to be effective at genome-editing in eukarytoic cells. We can also delivery SaCas9 in a much more

compact vector than by using standard promoters. **FIG. 22F** shows that the present invention is able to potential for far-far greater number of genomic targets, and hence a far greater number of potential mutations and diseases. **FIG. 22G** depicts an illustration of AAV virus size using the H1 bidirectional promoter to express SaCas9 and a single gRNA. Approximately 1kb of space is available as an HDR template. **FIG. 22H** depicts an illustration of AAV virus size using the H1 bidirectional promoter to express SaCas9 and a single gRNA and a template. **FIG. 22I** shows the following: 1. Adeno Associated Virus (AAV) is the safest and most-commonly used vector in gene therapy. The virus has one drawback: It is very small. 2. Naturally, it consists of a single-stranded DNA genome of 4700 nucleotides. The ends of the genome are inverted repeat sequences known as ITRs (orange). 3. These sequences (~150bp each) are the only required elements for packaging DNA into AAV, so the viral sequences (*Rep* and *Cap*) can be gutted. 4. Any cargo (blue) can be inserted in between the ITRs and the virus can be stuffed up to 5200 nucleotides total, meaning that the virus can accomodate cargo up to 4900 nucleotides. 5. Recombinant Adeno Associated Virus (rAAV) is used as a high-efficiency vehicle for delivering cargo into a cell. Once inside the cell the single-stranded genome will become double-stranded and will remain inside for the lifetime of the cell.

FIG. 23 shows the mechanism for site-specific recombination from a single AAV virus containing Cas9 a single gRNA and a template.

FIG. 24 comprises two panels, A and B. **FIG. 24A** shows potential configurations for HDR delivery within AAV vector. **FIG. 24B** shows potential configurations for HDR delivery within an intronic region of an RNA-directed polymerase within an AAV vector.

FIG. 25 comprises two panels, A and B. **FIG. 25A** shows an illustration of the Cas9 Nickase approach using the SaCas9 and a single gRNA and a template. **FIG. 25B** shows a depiction of the rd12 targeting sequence to correct a recessive RPE65 mutation.

FIG. 26 comprises two panels, A and B. **FIG. 26A** shows an illustration of AAV virus size using the H1 bidirectional promoter to express SpCas9 and a single gRNA. Approximately 0.7kb of space is available as an HDR template. **FIG. 26B** shows a generic approach for delivery of an RNA-directed nucleases, guide RNA, and template in a single AAV.

FIG. 27 depicts the rd10 targeting sequence to correct a recessive PDE6b mutation

FIG. 28 depicts a cloning vector that is easily customizable by insertion of HDR templates.

FIG. 29 contains two panels, A and B. **FIG. 29A** depicts a cartoon schematic and table of compact bidirectional promoter sequences with both pol II and III activity. **FIG. 29B** depicts a bar graph showing GFP expression of constructs comprising the compact bidirectional promoter sequences shown in **FIG. 29A**. **FIG. 29B** shows starting from the y-axis on the right and moving left along the x-axis, the following labels for the bars, 1) minus control, 2) hH1Koz (SEQ ID NO: 71 (or SEQ ID NO: 12 or 32)), 3) hSk1 (SEQ ID NO: 73), 4) biSRP (SEQ ID NO: 72), 5) biCGB1 (SEQ ID NO: 74), 6) biAloxE3 (SEQ ID NO: 77), 7) biDPP9 (SEQ ID NO: 79), 8) biTHEM (SEQ ID NO: 83), and 9) biORF (SEQ ID NO: 80).

FIG. 30 contains two panels, A and B. **FIG. 30A** depicts a cartoon schematic and table of orthologous bidirectional promoter sequences with both pol II and III activity. **FIG. 30B** depicts a bar graph showing GFP expression of constructs comprising the orthologous bidirectional promoter sequences shown in **FIG. 30A**. **FIG. 30B** shows starting from the y-axis on the right and moving left along the x-axis, the following labels for the bars, 1) minus control, 2) hH1+ (corresponding to human RPPH1-PARP2), 3) mH1+ (corresponding to mouse RPPH1-PARP2), and 4) rH1+ (corresponding to rat RPPH1-PARP2).

DETAILED DESCRIPTION

The presently disclosed subject matter now will be described more fully hereinafter with reference to the accompanying Figures, in which some, but not all embodiments of the presently disclosed subject matter are shown. Like numbers refer to like elements throughout. The presently disclosed subject matter may be embodied in many different forms and should not be construed as limited to the embodiments set forth herein; rather, these embodiments are provided so that this disclosure will satisfy applicable legal requirements. Indeed, many modifications and other embodiments of the presently disclosed subject matter set forth herein will come to mind to one skilled in the art to which the presently disclosed subject matter pertains having the benefit of the teachings presented in the foregoing descriptions and the associated Figures. Therefore, it is to be understood that the presently disclosed subject matter is not to be limited to the specific embodiments disclosed and that modifications and other embodiments are intended to be included within the scope of the appended claims.

Genome-editing technologies such as zinc fingers nucleases (ZFN) (Porteus, and Baltimore (2003) *Science* 300: 763; Miller et al. (2007) *Nat. Biotechnol.* 25:778-785; Sander et al. (2011) *Nature Methods* 8:67-69; Wood et al. (2011) *Science* 333:307) and transcription activator–like effectors nucleases (TALEN) (Wood et al. (2011) *Science* 333:307; Boch et al. (2009) *Science* 326:1509-1512; Moscou and Bogdanove (2009) *Science* 326:1501; Christian et al. (2010) *Genetics* 186:757-761; Miller et al. (2011) *Nat. Biotechnol.* 29:143-148; Zhang et al. (2011) *Nat. Biotechnol.* 29:149-153; Reyon et al. (2012) *Nat. Biotechnol.* 30:460-465) have empowered the ability to generate targeted genome modifications and offer the potential to correct disease mutations with precision. While effective, these technologies are encumbered by practical limitations as both ZFN and TALEN pairs require synthesizing large and unique recognition proteins for a given DNA target site. Several groups have recently reported high-efficiency genome editing through the use of an engineered type II CRISPR/Cas9 system that circumvents these key limitations (Cong et al. (2013) *Science* 339:819-823; Jinek et al. (2013) *eLife* 2:e00471; Mali et al. (2013) *Science* 339:823-826; Cho et al. (2013) *Nat. Biotechnol.* 31:230-232; Hwang et al. (2013) *Nat. Biotechnol.* 31:227-229). Unlike ZFNs and TALENs, which are relatively time consuming and arduous to make, the CRISPR constructs, which rely upon the nuclease activity of the Cas9 protein coupled with a synthetic guide RNA (gRNA), are simple and fast to synthesize and can be multiplexed. However, despite the relative ease of their synthesis, CRISPRs have technological restrictions related to their access to targetable genome space, which is a function of both the properties of Cas9 itself and the synthesis of its gRNA.

Cleavage by the CRISPR system requires complementary base pairing of the gRNA to a 20-nucleotide DNA sequence and the requisite protospacer-adjacent motif (PAM), a short nucleotide motif found 3' to the target site (Jinek et al. (2012) *Science* 337: 816-821). One can, theoretically, target any unique N_{20}-PAM sequence in the genome using CRISPR technology. The DNA binding specificity of the PAM sequence, which varies depending upon the species of origin of the specific Cas9 employed, provides one constraint. Currently, the least restrictive and most commonly used Cas9 protein is from *S. pyogenes*, which recognizes the sequence NGG, and thus, any unique 21-nucleotide sequence in the genome followed by two guanosine nucleotides (N_{20}NGG) can be targeted. Expansion of the available targeting space imposed by the protein component is limited to the discovery and use of novel Cas9 proteins with altered PAM requirements (Cong et al. (2013) *Science*

339: 819-823; Hou et al. (2013) *Proc. Natl. Acad. Sci. U.S.A.*, 110(39):15644-9), or pending the generation of novel Cas9 variants via mutagenesis or directed evolution. The second technological constraint of the CRISPR system arises from gRNA expression initiating at a 5' guanosine nucleotide. Use of the type III class of RNA polymerase III promoters has been particularly amenable for gRNA expression because these short non-coding transcripts have well-defined ends, and all the necessary elements for transcription, with the exclusion of the 1+ nucleotide, are contained in the upstream promoter region. However, since the commonly used U6 promoter requires a guanosine nucleotide to initiate transcription, use of the U6 promoter has further constrained genomic targeting sites to $GN_{19}NGG$ (Mali et al. (2013) *Science* 339:823-826; Ding et al. (2013) *Cell Stem Cell* 12:393-394). Alternative approaches, such as *in vitro* transcription by T7, T3, or SP6 promoters, would also require initiating guanosine nucleotide(s) (Adhya et al. (1981) *Proc. Natl. Acad. Sci. U.S.A.* 78:147-151; Melton et al. (1984) *Nucleic Acids Res.* 12:7035-7056; Pleiss et al. (1998) *RNA* 4:1313-1317).

The presently disclosed subject matter relates to compositions and methods comprising improvements of a CRISPR/Cas9 system (i.e., CRISPR guide RNAs using the H1 promoter; WO2015/19561, herein incorporated by reference in its entirety). Such a modified CRISPR/Cas9 system may comprise modifications to the H1 promoter region. In some embodiments, the modified CRISPR/Cas9 system comprises enhancing H1 bidirectional pol II expression using 5'UTR modifications. In some embodiments, the modified CRISPR/Cas9 system comprises modulating bidirectional expression through use of different orthologous sequences of the H1 promoter. In some embodiments, the modified CRISPR/Cas9 system comprises novel compact bidirectional promoter (including compact and orthologous promoter) sequences with both pol II and pol III activity (e.g., 7sk, 5'UTRs, Kozak consensus sequences, or combinations thereof).

In some embodiments, the modified CRISPR/Cas9 system comprises conditional pol II/pol III bidirectional promoter (including compact and orthologous promoter) expression (e.g., TetR and TetO sites) which can regulate ribonucleoprotein enzymatic activity or RNA-directed nucleases. In some embodiments, the improvements comprise addition of a donor template sequence for correcting mutations (e.g., homology directed repair (HDR)).

In some embodiments, compact bidirectional promoters include, but not limited to, RPPH1-PARP2 (Human), SRP-RPS29, 7sk1-GSTA4, SNAR-G-1-CGB1, SNAR-CGB2,

RMRP-CCDC107, tRNA(Lys)-ALOXE3, RNU6-9-MED16: tRNA (Gly)-DPP9, RNU6-2-THEM259, or SNORD13-C8orf41.

In some embodiments, orthologous bidirectional promoters include, but not limited to, RPPH1-PARP2 (Mouse) or RPPH1-PARP2 (Rat), or those derived from *ailuropoda melanoleuca, bos taurus, callithrix jacchus, canis familiaris, cavia porcellus, chlorocebus sabaeus, choloepus hoffmanni, dasypus novemcinctus, dipodomys ordii, equus caballus, erinaceus europaeus, felis catus, gorilla gorilla, homo sapiens, ictidomys tridecemlineatus, loxodonta africana, macaca mulatta, mus musculus, mustela putorius furo, myotis lucifugus, nomascus leucogenys, ochotona princeps, oryctolagus cuniculus, otolemur garnettii, ovis aries, pan troglodytes, papio anubis, pongo abelii, procavia capensis, pteropus vampyrus, rattus norvegicus, sus scrofa, tarsius syrichta, tupaia belangeri, tursiops truncatus, vicugna pacos*.

Table 3: Examples of compact bidirectional promoters

ncRNA gene	Protein gene	Distance
RPPH1	PARP2	230bp
SRP	RPS29	233bp
7sk1	GSTA4	239bp
SNAR-G1	CGB1	308bp
SNAR	CGB2	308bp
RMRP	CCDC107	361bp
tRNA(Lys)	ALOXE3	376bp
RNU6-9	MED16	412bp
tRNA (Gly)	DPP9	484bp
SNORD13	C8orf41	847bp
RNU6-2	THEM259	525bp

RPPH1-PARP2 (Human):
GGAATTCGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAGTG
TCACTAGGCGGGAACACCCAGCGCGCGTGCGCCCTGGCAGGAAGATGGCTGTG
AGGGACAGGGGAGTGGCGCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGA
AATCACCATAAACGTGAAATGTCTTTGGATTTGGGAATCTTATAAGTTCTGTAT
GAGACCACTTTTTCCC (SEQ ID NO: 71)

SRP-RPS29:
CTTGCTCTCAGCAGTGCAACGAGGTAAAAGGAAGAAGCTGGCCCACGCATGCG
CTCTTCAAATTTTTGAGACAGTTTACCCAGAATGCAGTGCTCAAAGGAAACGCG
TGCGCAGTGTGGTCAGGTTGTTTCGCTGGGTGAGTAAAATGAAATCTTAGAGGC
GTTGTGGGCTGGCCCAGTTGATGACGTCACCATACCACAGCTTCTAGTGCTATT
CTGCGCCGGTATCCGACC (SEQ ID NO: 72)

7sk1-GSTA4:
AGTATTTAGCATGCCCCACCCATCTGCAAGGCATTCTGGATAGTGTCAAAACAG
CCGGAAATCAAGTCCGTTTATCTCAAACTTTAGCATTTTGGGAATAAATGATAT
TTGCTATGCTGGTTAAATTAGATTTTAGTTAAATTTCCTGCTGAAGCTCTAGTAC
GATAAGCAACTTGACCTAAGTGTAAAGTTGAGACTTCCTTCAGGTTTATATAGC
TTGTGCGCCGCTTGGGTACCTC (SEQ ID NO: 73)

SNAR-G-1- CGB1:
GTCTCTCTCTTAGCGGGATATCTTCCGCAAGCACTGGGAATGTGGACATGGAAA
GTAAATTGAGTCTCCGTGGGGGAGTGAGACAGGGAGTGAGGGGTGTTGGACGC
GGCACGGGAACCTGGCCAGAGTCAGCGGACCCAATTGGCTGCTCTCTCTCAGA
TGCAGTTCCCCTTCCTCCCTCCAGGGGGCGCCACGGAACGCAGGGCCCTCACTG
GCCCTGGGGACTGGGTGACGTCAGGGATGAGCCTCTTGTGATTGGCTCCATCAC
CCTGCGTAAGATCAAAGGGAAGAAAGGATGGGCCCGACAA (SEQ ID NO: 74)

SNAR- CGB2:
GTCTCTCTCTTAGCGGGATATCTTCCGCAAGCACTGGGGATGTGGACATGGAAA
GTAAATTGAGTCTCCGTGGGGGAGTGAGACAGGGAGTGAGGGGTGTTGGACGC
GGCACGGGAACCCGGCCGGAGTCAGCGGACCCAATTGGCTGCTCTCTCTCAGA
TACAGTTCCCCTTCCTCCCTCCAGGGGGCGCCACGGAACGCAGGGCCCTCACTG
GCCCTGGGGACTGGGTGACGTCAGGGGTGAGCCTCTCCTGATTGGCTCCATCAC
CCTGCGTAAGGTCAAAAGGAAGAAAGGAGATCCCCGACAC (SEQ ID NO: 75)

RMRP-CCDC107:
TGCCGGCCCACGGGTGGAGGGATCGGGCGGGCGGTGCCGAAGCGGTCCGGCAT
TGGCCGGCCGCCCCAACGCGCACGCGCACGCGAGCAGGCCGGCCGGCTCCGGG

GAGGCCACGCCCACTCCCCGTAGGGCGGGGCCAGACCATATTTGCATAAGATA
GTGTCATTCTAGCTTTCCTGTATTTGTTCATTTCGTGTCTATTAGCTATTCTGCTA
GCCACAATGCCTCTGAAAGCCTATAGTCTTAGAAAGTTATGCCCGAAAACGGTT
TTTTTAATCTCACGCCACCAACTTTCTCACCCTAATCATAAAACACAATTTCTTT
AGGGCTATAAAATACTACTCTGTGAAGCTGAGGACGT (SEQ ID NO: 76)

tRNA(Lys)-ALOXE3:
TCTTTCCGCTCCAGGACCGCCCTGGGCCTGCAGGATCCTGGGCGGGAGCCCAGG
TGTCCGGGATCTGGGCCACTAGGGACTGGGGAGGAACCTCTCAGAGAAGCCCA
TAGCCCGCAGCGGCCCCGCGCGGCCGGTTCCGGCGCCGCACTGTTCCAGCCTCT
ACTATGGTACAGTCCCTGCGTCGCAGCCTCGGCGGGGGCTCTAAGAACGGGAG
GCAGAAAAAGCTCAATCAGCAGCAGGCGAGCTTCACCCGCTGCTTCCAAATCT
GTGCCAAAATATTCTATGCTGCACAGATAAAATCCTCTGTCGGTTCTACAAGCC
TGGCTTTTCCTATAGAGAACCCTCTTATAAGCAAAAAGTAAAGCTCTCGTGAAG
A (SEQ ID NO: 77)

RNU6-9-MED16:
GAGGGCAGTCACCAGCTCCTGGCCCGTGCGCCAAGCTCAGCGGGCGTCCGCGG
TGCGATCTTCCCTAGCGCCTCGGGTCTGGCGCCGCCATCTTCCTCGGTAACAAC
CAGTCGCCTGAGGCGTGGGGCCGCCTCCCAAAGACTTCTGGGAGGGCGGTGCG
GCTCAGGCTCTGCCCCGCCTCCGGGGCTATTTGCATACGACCATTTCCAGTAAT
TCCCAGCAGCCACCGTAGCTATATTTGGTAGAACAACGAGCACTTTCTCAACTC
CAGTCAATAACTACGTTAGTTGCATTACACATTGGGCTAATATAAATAGAGGTT
AAATCTCTAGGTCATTTAAGAGAAGTCGGCCTATGTGTACAGACATTTGTTCCA
GGGGCTTTAAATAGCTGGTGGTGGAACTCAATATTC (SEQ ID NO: 78)

tRNA (Gly)-DPP9:
TAACCGCTCAGCTGACCTCAGGAGGGCAGGGGTGCCTTCTAAAGGGTCCAGAG
AGCCTCCATTCCAGCTGCAGGCGTGGGACACAGACCGGGACGTGGGGCGGCGG
CCGGACTGGGCAGGTCGTCCCGGGTCCAGCGGCGCCTCACGGTCGCGGCTCCA
TGCCCGGGACTGCGACCCCGGAAGTGGCGGGAGCGGGGGACGACAGCCGCGG
CGGACACAGGGGACCCGCCGGCTCAGGCACCTTTGACCCGGAAGTTGAGCGAC
CCAGGCGGCGGCCTGGGATTGGACACCACCAGGCACGTACCAAGGCGTCCGCG

GCGCTTGGGGGGGAGCCCGCGGCGCGGCGGCCTAAGGTGCGTAACGCCCCATG
AACGACATCTTCCGGTGGGTTAGGGAGAGACACCCCCCTGTGACTTGGTATCAC
TCAGTCAAACCCATGATCCCCCACTATTAAGGATATCCGGAGAGGATGCTACCT
ATCAGG (SEQ ID NO: 79)

SNORD13- C8orf41:
TCCTGACTGCAGCACCAGAAGGCTGGTCTCTCCCACAGAACGAGGATGGAGGC
GGGGAGGGATCCGTTGAAGAGGGAAGGAGCGATCACCCAAAGAGAACTAAAA
TCAAATAAAATAAAACAGAGAGATGTCTTGGAGGAGGGGGCGAGTCTGACCGG
GATAAGAATAAAGAGAAAGGGTGAACCCGGGAGGCGGAGTTTGCAGTGAGCC
GAGATCGCGCCACTGCACTCCAGCCTGGGCGACAGAGTGAGACTCCGTCTCAG
TAAAAAAAAAAAAAAAAAAAAGAATAAAGAGGAAAGGACGCAAGAAAGGGA
AAGGGGACTCTCAGGGAGTAAAAGAGTCTTACACTTTTAACAGTGACGTTAAA
AGACTACTGTTGCCTTTCTGAAGACTAAAAAGAAAAAAAACTTAAAAATTTAA
AGAAATAAACTTCTGAGCCATGTCACCAACTTAACCACCCCCAGGTACCTGCAA
CGGCTCGCGCCCGCCGGTGTCTAACAGGATCCGGACCTAGCTCATATTGCTGCC
GCAAAACGCAAGGCTAGCTTCCGCCAGTACTGCCGCAACACCTTCTTATTTCAC
GACGTATGGTCGTAAAGCAATAAAGATCCAGGCTCGGGAAAATGACGGAGAG
GTGGAACTATAGAGAATAAATTTGCATATATAATAATCCGCTCGCTAATTGTGT
TTCTGTTTTCCTTTGCTAAGGTAGAAACAAAAGAATAATCACAGAATCTCAGTG
GGACTTTGAAAATATCCAGGATTTTATACGTGAAGAATGGATGTATCGCATTAC
GGTAGTCACCCTATGTGTAAATTAGTGGCACATACTTGGCACTCCTTAATGTCA
ACTATAAGATG (SEQ ID NO: 80)

RNU6-2-THEM259;
GCCTCCCAGCGTCGCGCCCTAACGACCCGCAAGTGTCCGAGGGCGCCTCCCGG
CCGCCATCGGCCGCCCTCGCAGCCGCCGCTCTCCTCACGGCCTCCCGGCCGCCG
CCGCCATCTTCCGCTTTCTCGTCCGGCTGCGGCGCTGCTGACGCTAGCGAGTCG
CCACGCCGGGCAAGAGCGGCCCCCCTGCGCCCGCAGAGAACGCTGGGATGCCA
GCGGCGCCCGCGGAGGCCTCACCCCCTACCTCGGCCGCTCCAGGGGGCGGGCC
TGCATCTGGGCCACCTCTTTTGCATATTGGCACCCACAATCCACCGCGGCTATG
AGGCCAGTATAAGGCGGTAAAATTACGATAAGATATGGGATTTTACGTGATCG
AAGACATCAAAGTAAGCGTAAGCACGAAAGTTGTTCTGCAACATACCACTGTA

GGAAATTATGCTAAATATGAAACCGACCATAAGTTATCCTAACCAAAAGATGA TTTGATTGAAGGGCTTAAAATAGGTGTGACAGTAACCCTTGAGTC (SEQ ID NO: 83)

Table 4: Examples of orthologous bidirectional promoters

ncRNA gene	Protein gene	Distance	Organism
RPPH1	PARP2	230bp	Human
RPPH1	PARP2	172bp	Mouse
RPPH1	PARP2	201bp	Rat

RPPH1-PARP2 (Mouse):
CGCTCTTGAAGGACGACGTCATCATCCCTTGCCCGGATGCGCGGGCTTCTTGTC TAGCACAGGAGCCTGGGGTAGAGCGCATGCAAATTACGCGCTGTGCTTTGTGG GAAATCACCCTAAACGAAAAATTTATTCCTCTTTCGAGCCTTATAGTGGCGGCC GGTCTACATCC (SEQ ID NO: 81)

RPPH1-PARP2 (Rat):
GGCTGATGAGCTTCCCCCGCCCACTAGGAGTGTGAAGACCTGCCGCCATAATA AGACTCCAAAAGACAGTGAATTTAACACTTACGGTGACTTCCCACAAAGCACA GCGTGTAATTTGCATGCGCTCTAGCCCAGGCTCCAGCTCCGGACCAGAAGCCCG CGCATCCCGGCAAAGGGTGATGACGTCGTCCTTCAAGCGCT (SEQ ID NO: 82)

In some embodiments, the modifications are made to nucleases (e.g., RNA-guided nucleases). In some embodiments, the nuclease (e.g., Cas9) is modified through post-transcriptional Cell-cycle regulation (e.g., fusion proteins comprising Geminin (Gem) or Cdt1). In some embodiments, the nuclease (e.g., Cas9) is modified by engineering partial target sites such that the nuclease can bind without DNA cleavage. In some embodiments, the nuclease (e.g., Cas9) is modified by modulating its half-life using N-terminal amino acid identity.

Such a modified CRISPR/Cas9 system can precisely target genomic sites, or facilitate the repair of a defective genomic sites, with greater efficacy, safety, and precision. Moreover, this modification provides a compact CRISPR/Cas9 system that allows for

higher-resolution targeting of oncogenes over existing CRISPR, TALEN, or Zinc-finger technologies.

Other aspects of the invention relate to methods comprising expression screens for additional bidirectional promoters with both RNA pol II and RNA pol III activities.

I. IMPROVEMENTS/MODIFICATIONS CRISPR/CAS9 SYSTEM.

A. Compositions

In some embodiments, the presently disclosed subject matter relates to compositions comprising improvements of a CRISPR/Cas9 system previously described in WO2015/195621 (herein incorporated by reference in its entirety). Such improvements comprise a non-naturally occurring nuclease system (e.g., CRISPR-Cas9) comprising one or more vectors comprising: a) a promoter (e.g., orthogolous H1 promoter or 7sk) operably linked to at least one nucleotide sequence encoding a nuclease system guide RNA (gRNA), wherein the gRNA hybridizes with a target sequence of a DNA molecule in a cell, and wherein the DNA molecule encodes one or more gene products expressed in the cell; and b) a regulatory element operable in a cell operably linked to a nucleotide sequence encoding a genome-targeted nuclease (e.g., modified Cas9 protein), wherein components (a) and (b) are located on the same or different vectors of the system, wherein the gRNA targets and hybridizes with the target sequence and the nuclease cleaves the DNA molecule to alter expression of the one or more gene products. In some embodiments, the system is packaged into a single adeno-associated virus (AAV) particle. In some embodiments, the system inactivates one or more gene products. In some embodiments, the system excises at least one gene mutation. In some embodiments, the H1 promoter comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 12 or 32. In some embodiments, the orthologous H1 promoter is derived from mouse or rat. In some embodiments, the orthologous H1 promoter comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NOs: 84-119. In some embodiments, the orthologous H1 promoter comprises a nucleotide sequences set forth in the group consisting of SEQ ID NOs: 84-119. In some embodiments, the H1 promoter comprises: a) control elements that provide for transcription in one direction of at least one nucleotide sequence encoding a gRNA; and b) control elements that provide for transcription in the opposite direction of a nucleotide sequence encoding a genome-targeted

nuclease. In some embodiments, the promoter is the human, mouse, or rat 7sk promoter. In some embodiments, the 7sk is selected from the group consisting of 7sk1 (variant 1 of 7sk which uses the endogenous GSTA4 5'UTR), 7sk2 (variant 2 of 7sk which uses just a kozak sequence in the 5'UTR), and 7sk3 (variant of 7sk which uses the beta globin 5'UTR). In some embodiments, the human 7sk1 comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 3. In some embodiments, the mouse 7sk1 comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to SEQ ID NO: 6. In some embodiments, the human 7sk2 comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to SEQ ID NO: 4. In some embodiments, the mouse 7sk2 comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to SEQ ID NO: 7. In some embodiments, the human 7sk3 comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to SEQ ID NO: 5. In some embodiments, the mouse 7sk3 comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to SEQ ID NO: 8. In some embodiments, the orthologous H1 promoter or 7sk promoter has at least one modification. In some embodiments, the at least one modification of the promoter comprises an element that allows conditional regulation. In some embodiments, the element is a tet-responsive promoter. In some embodiments, the tet-response promoter comprises a Tet repressor (TetR) and Tet operator (TetO) engineered into the H1 promoter. In some embodiments, the TetR comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to SEQ ID NO: 14. In some embodiments, the TetO comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 11. In some embodiments, the H1-TetO comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 13. In some embodiments, the at least one modification of the promoter comprises a site that allows auto-regulation. In some embodiments, the auto-regulation site comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 33 or 34. In some embodiments, the nuclease system further comprising at least one terminator sequence. In some embodiments, the terminator sequence is any nucleic acid selected from the group consisting of SEQ ID NOs: 120-130. In some embodiments, the at least one terminator

sequence is selected from the group consisting of SV40 or synthetic poly A (SPA) sequences. In some embodiments, the terminator sequences is a SV40 120 (SEQ ID NO: 123) or 240 (SEQ ID NO: 122) base pair sequence. In some embodiments, the terminator sequences is a SPA 49 base pair equence. In some embodiments, the nuclease system further comprises 5' untranslated region (5'UTR) sequences. In some embodiments, the nuclease system further comprises a Kozak sequence. In some embodiments, the Kozak sequence comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to SEQ ID NO: 1 or 2. In some embodiments, the nuclease system further comprises a RNA sequence that mediates cap-independent initiation of translation. In some embodiments, the RNA sequence is selected from the group consisting of 6.947 or 6.967 (Wellensiek I. (2013) *Nature Methods*, 10:747-750). In some embodiments, the Cas9 comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 61. In some embodiments, the Cas9 comprises at least one modification. In some embodiments, the at least one modification in the Cas9 comprises an alteration in the cleaving sequence. In some embodiments, the nuclease is seleted from the group consisting of Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 and Csx12), Cas10, Cas13a, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, Cpf1, C2c1, C2c2, and C2c3. In some embodiments, the alteration in the cleaving sequence selected from the group consisting of T2A, P2A, E2A, and F2A. In some embodiments, the T2A comprises an amino acid sequence. having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 36. In some embodiments, the P2A comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 35. In some embodiments, the P2A comprises an amino acid sequence having the nucleotide sequence set forth in SEQ ID NO: 35. In some embodiments, the E2A comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 37. In some embodiments, the F2A comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 38. In some embodiments, the at least one modification in the Cas9 comprises a codon optimized for

expression in the cell. In some embodiments, the Cas9 further comprises a linker sequence operably fused in frame to a cell-cycle dependent protein (Cas9 fusion). In some embodiments, the cell cycle-dependent protein is selected from the group consisting of APE2, ATR, 53BP1, BRCA1, Chk1, Cdc5, Cdc6, Cdc7, Cdc45, Cdt1, CSA, CSB, Ctf18, CtIP, DDB1, DDB2, DNA2, DUT, Elg1, EndoV, Esp1, Exonuclease1, FBH1, FEN1, Geminin, Hus1, KNTC2 (NDC80), Ku80, Ligase1, Mad2, MBD4, Mcm3, Mcm4, Mcm5, Mcm6, Mcm7, Mcm8, Mcm10, MGMT, MLH3, Mms4, MPG, MSH2, Mus81, NBS1, NEIL2, NEIL3, NTH1, Orc1, Orc3, PARP1, PCNA, Pif1, Pin1, PMS1, PMS2, PNK, Pola p180, Pola p70, Pola Spp1 (Prim2a), Polb, Pold p125, Pole Dpb3, Pole Dpb4, Pole Pol2, Poli, Poll, Polm, Psf1, Psf2, Psf3, Rad1, Rad18, Rad23A, Rad23B, Rad51, Rad51D, Rad54, Rad6A, Rag1, Rag2, RPA34, RPA70, Scc1, Scc3, Sir2, SIRT1 (Sirtuin), TDG, TDP1, TIMELESS, Tin2, Topoisomerase I, Topoisomerase IIIa, Topoisomerase IIIb, Ubc13, UNG, XAB2, XPC, XPF, XPG, Xrcc2, and XRCC4. In some embodiments, the cell cycle-dependent protein is Geminin. In some embodiments, the cell cycle-dependent protein is human Geminin. In some embodiments, the human Geminin comprises the amino acids from positions 1-110 (hGem(1-110)). In some embodiments, the hGem(1-110) comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the amimo acid sequence set forth in SEQ ID NO: 19. In some embodiments, the cell cycle-dependent protein is Cdt1. In some embodiments, the cell cycle-dependent protein is human Cdt1. In some embodiments, the human Cdt1 comprises amino acids from positions 30-120 (hCdt1(30-120)). In some embodiments, the hCdt1(30-120) comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the amino acid sequence set forth in SEQ ID NO: 18. In some embodiments, the Cas9 fusion comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the amino acid sequence selected from the group consisting of SEQ ID NO: 20-27. In some embodiments, the Cas9 is operably fused in frame to ubiquitin (Ub-Cas9). In some embodiments, the Ub-Cas9 at least one N-terminal modification. In some embodiments, the N-terminal modified Ub-Cas9 comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the amino acid sequence selected from the group consisting of SEQ ID NO: 39-58. In some embodiments, the ubiquitin protein comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 60. In some embodiments, the nuclease system further comprises a SaCas9 nickase. In some

embodiments, the nuclease system further comprises a donor template sequence. In some embodiments, the at least one vector comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 62 or 67. In some embodiments, the donor template sequence corrects at least one gene mutation. In some embodiments, the at least one gene mutation is *rd10* or *rd12*. In some embodiments, the at least one vector comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 63 or 64. In some embodiments, the at least one vector comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 65 or 66. In some embodiments, the promoter is operably linked to at least one, two, three, four, five, six, seven, eight, nine, or ten gRNA. In some embodiments, the target sequence comprises the nucleotide sequence $AN_{19}NGG$, $GN_{19}NGG$, $CN_{19}NGG$, or $TN_{19}NGG$. In some embodiments, the cell is a eukaryotic or or non-eukaryotic cell. In some embodiments, the eukaryotic cell is a mammalian or human cell. In some embodiments, the eukaryotic cell is a retinal photoreceptor cell. In some embodiments, the one or more gene products are rhodopsin. In some embodiments, the expression of the one or more gene products is decreased.

Table 5: Examples of Orthologous H1 sequences

mus_musculus

TTCAGGATGTAGACCGGCCGCCACTATAAGGCTCGAAAGAGGAATAAATTTTT CGTTTAGGGTGATTTCCCACAAAGCACAGCGCGTAATTTGCATGCGCTCTACCC CAGGCTCCTGTGCTAGACAAGAAGCCCGCGCATCCGGGCAAGGGATGATGACG TCGTCCTTCAAGAGCG (SEQ ID NO: 84)

rattus_norvegicus

AGGAGTGTGAAGACCTGCCGCCATAATAAGACTCCAAAAGACAGTGAATTTAA CACTTACGGTGACTTCCCACAAAGCACAGCGTGTAATTTGCATGCGCTCTAGCC CAGGCTCCAGCTCCGGACCAGAAGCCCGCGCATCCCGGCAAAGGGTGATGACG TCGTCCTTCAAGCGCT (SEQ ID NO: 85)

dipodomys_ordii

AGGAAAGACTTCGCTGAGGCAGACTTTATAAGGCTCCCGCGCAGAAAGAAACT TTATAGTTATGGTGATTTCCCACAAGCCACTGCGTCATGCAAATAAAGCAGGGT ACGGCTTCCATGTACCTTAAGGTTTTTTTCTAGGCCGCGTACGCTCTGCGTATTC

AGCCACGTGACCCTGAGCCAGTGGTTGTTGGGAGCACGTTGTGGACCTCTGCGT
TTGGATTCC (SEQ ID NO: 86)

ictidomys_tridecemlineatus

GAAAGGGACTCCGCACAAGCAGAGTTTATAAGGCTCCCATCTGTACAGCCATTT
CTCGGTCATGGTAACTACCCACAACACACAGCGATATGCAAATATAGCAGAGC
GTGTCTTCCCGCGCGCGCCTGGTCGTCTCGGCGCCGGCGCGCTGCGTGGGGCGG
AACTGTGACAGAGACCCTGCGATTCCTGGGAGCTGGCTGATGACATCAGTGTCT
AACCTCC (SEQ ID NO: 87)

cavia_porcellus

GAGAAAGAAAGGCTCAAACCTAGCCTTATAAGGCTCCCAAATGTCGGTATATT
TTTTGGTTATGGTGACTTCCCACAATGCATAGCGATATGTAGATATTGCCAGGA
GTACCTCCCACTTCTGGTCCTGTCAGCTCTTTTCTAGGACGCGCGCGCTGCAGGT
TTCCAGCCTGTGATTGGGCCAGCAATTCCGGGAATGAATTGATGACGTCAGCGT
TTGAATTCC (SEQ ID NO: 88)

ochotona_princeps

GGGGGAAGCTGGGCTCGATCAGCCTTTATAAAGCTCCAAAAACTCAAGACATT
TTTCTGTTACGGTGGCTTCCCACAGTACACAGCGACATGCAAATAGCTTGCCAA
TGAATTCGCGGACCGCTTCCCGCCCCGGCGCAGGCGCGCGGACGCTGTCTCCCC
TGGACGCGCGCTCGCGGTTCCCGGGAGCTGGCTGATGACGTTCGGTCTCC (SEQ
ID NO: 89)

oryctolagus_cuniculus

GGGGAGAGGTGGATCCGAACAGACTTTATAAAGCTCCGAAAGCCCAAGGCATC
TTTCCCTTACGGTAGCTTCCCACAAGACATAGCGACATGCAAATTTCTTGAAGT
ATGCTTCAGACGCGCTTCTCGCCACAGCGCAAGCGCGCTGTGTGCTGACGCGGG
AACGGGCCAGGGCGCGGTTCCCGGGAGCGGGTTGATGACGTTAGATCTCC (SEQ
ID NO: 90)

callithrix_jacchus

GAGGAAAAGTAGTCCCACAGACAACTTATAAGATTCCCATACCCTAAGACATT
TCACGATTATGGTGACTTCCCAGAAGACACAGCGACATGCAAATATTGCAGGT
CGTGTTTCGCCTGTCCCTCACAGTCGTCTTCCTGCCAGGGCGCACGCGCGCTGG
GTTTCCCGCCAACTGACGCTGGGCTCGCGATTCCTTGGAGCGGGTTGATGACGT
CAGCGTTTGAATTCC (SEQ ID NO: 91)

chlorocebus_sabaeus

GGGGAAGGGTGGTCCCTTACAGAACTTATAAGATTCCCAAACTCAAAGACATT
TCACGTTTATGGTGACTTCCCAGAAGACATAGCGACATGCAAATATTGCAGGGC
GTCACACCCCTCTCCCTCACAGTCATCTTCCTGCCAGGGCGCACGCGCGCTGGG
TGTTCTCGCGTAGTGACACTGGGCCCGCGATTCCTTGGAGCGGGTTGATGACGT
CAGCGTTCGAATTCC (SEQ ID NO: 92)

macaca_mulatta

GGGGAAGGGTGGTCCCACACAGAACTTATAAGATTCCCATACTCAAAGACATT
TCTCGTTTATGGTGACTTCCCAGAAGACACAGCGACATGCAAATATTGTAGGGC
GTCACACCCCTGTCCCTCACAGTCATCTTCCTGCCAGGGCGCACGCGCGCTGGG
TGTTCCCGCGTAGTGACACTGGGCCCGCGATTCCTTGGAGCGGGTTGATGACGT
CAGCGTTCGAATTCC (SEQ ID NO: 93)

papio_anubis

GGGGAAAGGTGGTACCATACAGAACTTATAAGATTCCCATACTCAAAGACATT
TCACGATTATGGTGACTTCCCAGAAGACACAGCGACATGCAAATATTGTAGGG
CGTCACACCCCCTGTCCCTCACAGTCATCTTCCTGCCAGGGCGCACGCGCGCTG
GGTGTTCCCGCGTAGTGACACTGGGCCCGCGATTCCTTGGAGCGGGTTGATGAC
GTCAGCGTTCGAATTCC (SEQ ID NO: 94)

gorilla_gorilla

GGGAAAGGGTGGTCCCACACAGAACTTATAAGACTCCCATATCCAAAGACATT
TCACGGTTATGGTGATTTCCCAGAACACATAGCGACATGTAAATATTGCAGGGC
GCCACTCCCCAGTCCCTCACAGCCATCTTCCTGCCAGGGCGCACGCGCGCTGGG
TGTTCCCGCCTAGTGACACTGGGCCCGCGATTCCTTGGAGCGGGTTGATGACGT
CAGCGTTCGAATTCC (SEQ ID NO: 95)

homo_sapiens

GGGAAAAAGTGGTCTCATACAGAACTTATAAGATTCCCAAATCCAAAGACATT
TCACGTTTATGGTGATTTCCCAGAACACATAGCGACATGCAAATATTGCAGGGC
GCCACTCCCCTGTCCCTCACAGCCATCTTCCTGCCAGGGCGCACGCGCGCTGGG
TGTTCCCGCCTAGTGACACTGGGCCCGCGATTCCTTGGAGCGGGTTGATGACGT
CAGCGTTCGAATTCC (SEQ ID NO: 96)

pan_troglodytes

GGGAAAGGGTGGTGCCACACAGAACTTATAAGATTCCCATATGCAAAGACATT
TCACGTTTATGGTGATTTCCCAGAACACATAGCGACATGCAAATATTGCAGGGC
GCCACTCCCCTGTCCCTCACTGCCATCTTCCTGCCAGGGCGCACGCGCGCTGGG

TGTTCCCGCCTAGTGACACTGGGCCCGCGATTCCTTGGAGCGGGTTGATGACGT
CAGCGTTCGAATTCC (SEQ ID NO: 97)

pongo_abelii

GAGAAAGGGTGGTCCCGTCCAGAACTTATAAGATTCCCATACCCAAAGACATT
TCACGTTTATGGTGACTTCCCAGAATGCATAGCGACATGCAAATATTGCAGGGC
GTCACTCCCCTGTCCCTCACAGCCATCTTCCTGCCAGGGCGCCCGCGCGCTGGT
GTTCCCGCCTAGTGACACTGGGCCCACGATTCCTTGGAGCGGGTTGATGACGTC
AGCGCTCGTATTCC (SEQ ID NO: 98)

nomascus_leucogenys

GGGGAAAAGTAGTAGACCTTATAAGATTCCCAAACCCAAAGACATTTCTCGTTT
ATGGTGACTTCCCAGAAGACATAGCGACATGCAAATATTGCAGGGCGCCACTC
CCCTGTCCCTCACAGCCATCTTCCTGCCAGGGCGCACGCGCGCTGGGTGTTCCC
GCCTAGTGACACTCGGCCCGCGATTCCTTGGAGCGGGTTGATGACGTCAGCGTT
CGAATTCC (SEQ ID NO: 99)

tarsius_syrichta

GCGAGAGGGTGGGTCCACACAGAGCTTATAAGGTTCACAAGTAAAGATATTTC
ACGGTGACGGTGACTTCCCACAATACACTGCGACATGCAAATATAGCCGGGCG
TGCCTCCCCGATCCCGGAAGAGCGACTCCTAGCCAGTGCGCACGCGCGCTGCG
TGTTCGCGTCCTAGGTCGCTGGGCCCGCGGTTCCTGGGAGCGGGTGGTGACGTC
AGCGGCCCAGCTTC (SEQ ID NO: 100)

otolemur_garnettii

GCCTAAAAGGGCGCTTGCACAGAATTTATAAGGTTCCCAAACAGAGACACATT
TCATTATTATGGTGACTTCCCACAATGCACAGCGCCATGCAAATATGCTAGGAC
GCCTCCCCCCGCTACCTTAAGGTCGTCAACTAACCAGTGCGCGCGCGCACTGCG
CGTTTCCCGCCGGTGACTCAATGCCCGCGTTTGGTGGGAGCTAGTTGGTGACCT
CAGTTCTGGAGGCTC (SEQ ID NO: 101)

tupaia_belangeri

GGGGGAAGCTGGGTCCACTGAGTTCTTATAAGGTTTCCAGTCCTAGAGCGATTT
TACCATTGCGGTGATTTCCCAGCATCCGTAGCTACATGCAAATAGCGCGGGGCG
CGTCTCTCAGGTCCCTCCCCGCCCTCTCACTGTACGTACCCGCGTCCTAGGGAC
GCCGCGCCCGGGGTTCCCGGACGTCAGCGTTCCGACGCA (SEQ ID NO: 102)

ailuropoda_melanoleuca

AGGGAAAGCCGCGCCTGGGGCGGATTTATAAGGCTTCCATATCTAAAGGCATT
TCACAGTCATGGTGACTTCCCACAATACATAGCAACATGCAAATATCGCGGGG
AGAACCTCCCCTGTCCCTTGTACGCGGCTTCTAAAGACGCACGCGCGCTCTGTG
TTCCCGCCCTGTGACTCTAGGCGGGCAATTCCTGGGACAGTGTTCTGACGGGAA
CGTTCAGGCTCC (SEQ ID NO: 103)

mustela_putorius_furo

GGGAAAGGGTGGACCCACCGAGCATTTATAAGGCTCCCGCATCTAAAGACATT
TTACAGTTATGGTGACTTCCCACAACGCGTAGCAACATGCAAATATCGTGGAGA
GTACCGCCCCTGTCCCATGCACGCGTCTTCTCAGCAGCACGCACGCGCGCTGTG
TTCCCGCCCTGTGACTCCAGGCGGGTATTTCCAGGGGCGGGTTGCTGACAGGAA
CGTTCAGGCTTC (SEQ ID NO: 104)

canis_familiaris

GCAGCGCAGCCCTCTCGCCGCTTATAAAGTGCCGCCCGCACGGCCCTTCTCGCT
CACGGCGACTTCCCATAACACACAGCAGCATGCAAATACCGCGGGGAGCCCCG
CCCCGCCCCGGCCCCCGCACCGCCTCGGGACGCATGCGCCGGCTCTCCGTTCCC
GCCTTGGGCCGGCGGCGGGCGGGCGGGCGAGCGGGCGGGAGCGGCTCCGGCG
GGGACGAGCGGGCGCC (SEQ ID NO: 105)

felis_catus

GGGAAAGGGTGGCCCCGCCGAGCATTTATAAGACTCCCATACCTAAAGACATT
TCTCAGTTATGGTGATTTCCCACAACACACAGCAACATGCAAATATCGAGGGGT
GTACCGCCCCTGTCCTTTGTAGACGTCTTCTCTCCAGGACGCACGCGCGCTGTA
TTCCCGCCTTGTGACTCTAGGCGGGCGATTCCTGGGAGAGGGTTGATGACGTCC
AAGTTCTGGCTTC (SEQ ID NO: 106)

equus_caballus

GGGGGAAAACAGCCCATGGCTGCATTTATAAGACTCACAGATCTAAAGCCATT
TCACGAATAGGGTGACTTCCCACAATACACAGCGACATGCAAACATAGCGGGG
CGTGCCTTTCCTGTACCCTGTGGGCATCTCTCCTGGACGCACGCGCGCCGGGTG
TTCCCGCGCTGTGACTCTAGGCAAGCGCTTCCTGGGAGAGAGTTGATGACGGCA
GCATTCGGGCTCC (SEQ ID NO: 107)

myotis_lucifugus

GGGAGAAGGAGGCGTAGAGGATATATAAGGCCCCCTTATGTGTAGTCCTTTTA
CGGTTAGGGTGACTTCCCACAACGCATAGCGACATGCAAATTTGACGGGCGTG
CCTCCTCTGTCCCTGCGGGCAACTTCTCTCCTGGACGCGCGCGCGCTGCGTGTTC

CCGCCTTTTGACTCCAGCCGAGCGAATCCTGGGAGAGGGCAGGTGACGTCAAC
AGTCAGGCTCG (SEQ ID NO: 108)

pteropus_vampyrus

GCGAGAAAAATTCTTCACGCAGAATATATAAGGATCCCATATCTGAAGACATTT
TACGATTACGGCGATTTCCCACAACACATAGCGACATGTAAATGTAGTGGGGC
ATGCCTCCCCTGTCCCTTGTGGGCAGCTTCTCGCCAGAACGCACGCGCGGTGCG
TGTTCCCGCCTTGTGACTAAGTTGGCGAGTCAGGGAGGAGATTGATGACGTCAG
CTCACCCGCTCC (SEQ ID NO: 109)

bos_taurus

GGCAAACACCGCACGCAAATAGCACTTATAATGTGCTCATACCTAGAGCCACTT
TTCGGTTACGGTGACTTCTCAAAAAGACAGTGGAACATGCAAATATTACAGTGC
GTCCCGCCCCTGGTAGGTCTACGCTAGGACGCACGCGCACTACGGTTCCCGCCT
ATAGACTGCGCTGGCGATTCCTGGGAGCGGACTGATGACGTCAGCGTTCGGGA
TCC (SEQ ID NO: 110)

ovis_aries

GGCGAACAATGCGCGCAAACAGCATTTATAATGAGCTCATACCTAAAGCCACT
TTACGGTTACGGTGACTTCCCACAAGACATTGCGGCATGCAAATATTTTAGTGC
GTCCCGCCCCTGGTAGTTCCACGCTAGGACGCACACGCACTACGGTTCCCGCCT
TTAGACTGCGCTGGCGATTCCAGGAGCGGACTGATGACGTCAGCGTTGGGGCT
CC (SEQ ID NO: 111)

tursiops_truncatus

GCCGAAAACCAGGCTCAAGCCACATTTATAAGGCTCCCAAATCTAAGTACATTT
GTCGGTTATGGTGACTTCCCGCACCACATTGCGACATGCAAATACTGCGGAGCG
TCCCTCCCCTGGCAACTCCTCGCTGGGACGCACGCGCGCTACGTGCTCCCGCCT
TTTGACTGCGCCGGCGATACTTGGGAGAGGGTTGATGACGTCAGCGTTCTGGCT
CC (SEQ ID NO: 112)

vicugna_pacos

GGGAAAGGGTGGGCTCACGCAGCCTTTATAAGACTCCCAAACTTAAAGACATT
TCTCGGTTATGGCGACTTCCCACAAGACATAGCGACATGCAAATACTGCAGGG
CGCCGACCCGGTCCTGTGCAGCCATCTTTCGGCTGGGACGCACGCGCGCTGCGT
GTTCCCGCCCTGTGACTGCGCCGGCGATTACTGGGAGAGGATTGATGACGTCAA
CGTTCGGGTTCC (SEQ ID NO: 113)

sus_scrofa

GTAGGAAAACTGCTTCTGTGAGCACTTATAAAACTCCCATAAGTAGAGAGATTT
CATAGTTATGGTGATTTCCCATAAGACATTGCGACATGCAAATATTGTGGCGCG
TTCGTCCCCGTCCGGTGCAGGCAGCTTCGCTCCAGGACGCACGCGCAATACATG
TTCCCGCCTTGAGACTGCGCCGGCAGATTCCTAGGAAGTGGTTGATGACGTCGA
TGTTAGGGATCC (SEQ ID NO: 114)

erinaceus_europaeus

GCCTAAACCGGCTCTTTCGACAGACTTATAAGGACCTCTTATCTTAGGACATTT
TTTTGTTAGGGTAACTTCCCACGATGCATAGCGATATGTAAATATGGCGCCGCG
AGTCTCTCCTAGGCGTCTCCCCAGGACGCAGGCGCACTGCTTGTTCCCGCGTTA
ACATTGCTGATTCTGGGAGACTGCTGATGACGTCAGCGTCCAGTCTAC (SEQ ID
NO: 115)

choloepus_hoffmanni

AGAAAAAAATAGTTTATGCTGGATTTATAAGATTCCCAAATCTAAAGCCATTTC
ACAGTTACGGTGATTCCCCACTACACACGGCGATATGCAAATATAGCGGAAGT
GTTCCTGAGGCGTGGTAAAGCGCGCGCGCGCTGAGAGTTCCCGCCCTGTGGTGC
TGGGCTGGAGATGCCTGAGAACTGGCTGATGACGGCAACGTTCGGGCTCC (SEQ
ID NO: 116)

dasypus_novemcinctus

AAAGCGATAGTTTTTTAAACTGGACTTATAAGGCACCCATATCTACGTATATTT
CATGGTTAGGGTGATTTCCCACAACACATAGCGAAATGCAAATATTGGAGGGC
GCTGAGGCGTGGTCGGGCGCAAGCGCGCTGCGACTTCCCGCCTTTCGGCCCTAG
GCCCCAGATTCCTGGGAGCTGGATGATGACGTTGACGTTCGGATACC (SEQ ID
NO: 117)

loxodonta_africana

GGGAAGGAACAAATTCGTCAGGATTTATAAGACTCTCAGAGCTGTAGACATTT
CACAGTTAGGGCGATGTCCCACAATACATAGCAACATGCAAATATTCTAGGAG
GCCAGCCTCCCCGTCCGCGTGGTCATCTTCTCGCTAGGGCGCACGCCCGCTGCG
TGTTCCCGCTCTGTGACCAGGCAGGCGATTCCTGAGAACCGCTTGGTGACGTCA
GTGTTCTGGCTCC (SEQ ID NO: 118)

procavia_capensis

AGGGTAAATCGGCGCTGCTCAGCATTTAAAAGAATCCCAAATGTGTCGC
CATTTTACGCTTAGGGTGATATCCCACAAGACACAGCGACATGCAAATATCGTG
AGTCTCTGTTTCCCTGTCCACGAGGGCGTCCTCTCGCTGGGGCGCACGCGCGGT

GTGTGTGCCCCCGTTGTGTGTTCCCGCGATTCCAAAGAACTGGTTGATAACGTT AGACTTCCGGCTGC (SEQ ID NO: 119)

Table 6: Examples of terminators

SPA

AATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTG (SEQ ID NO: 120)

SPA and Pause

AATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGA TAGTACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAAACTAGCA AAATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCT (SEQ ID NO: 121)

SV40 (240bp)

ATCTAGATAACTGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGC TTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAAT TGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAG CATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTG TCCAAACTCATCAATGTATCTTA (SEQ ID NO: 122)

SV40-mini (120bp)

TTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTC ACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCA ATGTATCTTAT (SEQ ID NO: 123)

bGH polyA

CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTC CTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAAT TGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCA GGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCG GTGGGCTCTATGG (SEQ ID NO: 124)

TK polyA

GGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGC TATGACGGCAATAAAAAGACAGAATAAAACGCACGGGTGTTGGGTCGTTTGTT CATAAACGCGGGGTTCGGTCCCAGGGCTGGCACTCTGTCGATACCCCACCGAG ACCCCATTGGGGCCAATACGCCCGCGTTTCTTCCTTTTCCCCACCCCACCCCCCA

AGTTCGGGTGAAGGCCCAGGGCTCGCAGCCAACGTCGGGGCGGCAGGCCCTGC CATAG (SEQ ID NO: 125)

sNRP1

GGTATCAAATAAAATACGAAATGTGACAGATT (SEQ ID NO: 126)

sNRP1a

AAATAAAATACGAAATGTGACAGATT (SEQ ID NO: 127)

Histone H4B

GGTTGCTGATTTCTCCACAGCTTGCATTTCTGAACCAAAGGCCCTTTTCAGGGC CGCCCAACTAAACAAAAGAAGAGCTGTATCCATTAAGTCAAGAAGC (SEQ ID NO: 128)

MALAT-1

GATTCGTCAGTAGGGTTGTAAAGGTTTTTCTTTTCCTGAGAAAACAACCTTTTGT TTTCTCAGGTTTTGCTTTTTGGCCTTTCCCTAGCTTTAAAAAAAAAAAAGCAAA AGACGCTGGTGGCTGGCACTCCTGGTTTCCAGGACGGGGTTCAAGTCCCTGCGG TGTCTTTGCTT (SEQ ID NO: 129)

MALAT-comp14

AAAGGTTTTTCTTTTCCTGAGAAATTTCTCAGGTTTTGCTTTTTAAAAAAAAAGC AAAAGACGCTGGTGGCTGGCACTCCTGGTTTCCAGGACGGGGTTCAAGTCCCTG CGGTGTCTTTGCTT (SEQ ID NO: 130)

The term “homologous” refers to the "% homology" and is used interchangeably herein with the term "% identity" herein, and relates to the level of nucleic acid sequence identity when aligned using a sequence alignment program.

For example, as used herein, 80% homology means the same thing as 80% sequence identity determined by a defined algorithm, and accordingly a homologue of a given sequence has greater than 80% sequence identity over a length of the given sequence. Preferred levels of sequence identity include, but are not limited to about, 60%, 65%, 70%,75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8% or more sequence identity to the nucleotide sequences set forth in SEQ ID NOs: 1-82. Exemplary levels of sequence identity include, but are not limited to about, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8% or more sequence identity to the nucleotide sequences set forth in SEQ ID NO: 1-82.

In some embodiments, the presently disclosed subject matter provides a non-naturally occurring CRISPR-Cas system comprising one or more vectors comprising: a) an H1 promoter operably linked to at least one nucleotide sequence encoding a CRISPR-Cas system guide RNA (gRNA), wherein the gRNA hybridizes with a target sequence of a DNA molecule in a cell, and wherein the DNA molecule encodes one or more gene products expressed in the cell; and b) a regulatory element operable in a cell operably linked to a nucleotide sequence encoding a Cas9 protein, wherein components (a) and (b) are located on the same or different vectors of the system, wherein the gRNA targets and hybridizes with the target sequence and the Cas9 protein cleaves the DNA molecule to alter expression of the one or more gene products.

In some embodiments, the presently disclosed subject matter provides a non-naturally occurring CRISPR-Cas system comprising one or more vectors comprising: a) an H1 promoter operably linked to at least one nucleotide sequence encoding a CRISPR-Cas system guide RNA (gRNA), wherein the gRNA hybridizes with a target sequence of a DNA molecule in a eukaryotic cell, and wherein the DNA molecule encodes one or more gene products expressed in the eukaryotic cell; and b) a regulatory element operable in a eukaryotic cell operably linked to a nucleotide sequence encoding a Type-II Cas9 protein, wherein components (a) and (b) are located on the same or different vectors of the system, whereby the gRNA targets and hybridizes with the target sequence and the Cas9 protein cleaves the DNA molecule, and whereby expression of the one or more gene products is altered. In one aspect, the target sequence can be a target sequence that starts with any nucleotide, for example, $N_{20}NGG$. In some embodiments, the target sequence comprises the nucleotide sequence $AN_{19}NGG$. In some embodiments, the target sequence comprises the nucleotide sequence $GN_{19}NGG$. In some embodiments, the target sequence comprises the nucleotide sequence $CN_{19}NGG$. In some embodiments, the target sequence comprises the nucleotide sequence $TN_{19}NGG$. In some embodiments, the target sequence comprises the nucleotide sequence $AN_{19}NGG$ or $GN_{19}NGG$. In another aspect, the Cas9 protein is codon optimized for expression in the cell. In another aspect, the Cas9 protein is codon optimized for expression in the eukaryotic cell. In a further aspect, the eukaryotic cell is a mammalian or human cell. In yet another aspect, the expression of the one or more gene products is decreased.

The presently disclosed subject matter also provides a non-naturally occurring CRISPR-Cas system comprising a vector comprising a bidirectional H1 promoter, wherein

the bidirectional H1 promoter comprises: a) control elements that provide for transcription in one direction of at least one nucleotide sequence encoding a CRISPR-Cas system guide RNA (gRNA), wherein the gRNA hybridizes with a target sequence of a DNA molecule in a eukaryotic cell, and wherein the DNA molecule encodes one or more gene products expressed in the eukaryotic cell; and b) control elements that provide for transcription in the opposite direction of a nucleotide sequence encoding a Type-II Cas9 protein, whereby the gRNA targets and hybridizes with the target sequence and the Cas9 protein cleaves the DNA molecule, and whereby expression of the one or more gene products is altered. In one aspect, the target sequence can be a target sequence that starts with any nucleotide, for example, N_{20}NGG. In some embodiments, the target sequence comprises the nucleotide sequence AN_{19}NGG. In some embodiments, the target sequence comprises the nucleotide sequence GN_{19}NGG. In some embodiments, the target sequence comprises the nucleotide sequence CN_{19}NGG. In some embodiments, the target sequence comprises the nucleotide sequence TN_{19}NGG. In some embodiments, the target sequence comprises the nucleotide sequence AN_{19}NGG or GN_{19}NGG. In another aspect, the Cas9 protein is codon optimized for expression in the cell. In another aspect, the Cas9 protein is codon optimized for expression in the eukaryotic cell. In a further aspect, the eukaryotic cell is a mammalian or human cell. In yet another aspect, the expression of the one or more gene products is decreased.

In some embodiments, the CRISPR complex comprises one or more nuclear localization sequences of sufficient strength to drive accumulation of the CRISPR complex in a detectable amount in the nucleus of a cell (e.g., eukaryotic cell). Without wishing to be bound by theory, it is believed that a nuclear localization sequence is not necessary for CRISPR complex activity in eukaryotes, but that including such sequences enhances activity of the system, especially as to targeting nucleic acid molecules in the nucleus. In some embodiments, the CRISPR enzyme is a type II CRISPR system enzyme. In some embodiments, the CRISPR enzyme is a Cas9 enzyme. In some embodiments, the Cas9 enzyme is *C. jejuni*, *G. stearothermophilus*, *N. meningitidis*, *S. aureus*, *S. pneumoniae*, *S. pyogenes*, or *S. thermophilus* Cas9, and may include mutated Cas9 derived from these organisms. The enzyme may be a Cas9 homolog or ortholog such as Cpf1 or Cas13a/C2c2.

In general, and throughout this specification, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. Vectors include, but are not limited to, nucleic acid molecules that are single-stranded,

double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g. circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. One type of vector is a "plasmid," which refers to a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Another type of vector is a viral vector, wherein virally-derived DNA or RNA sequences are present in the vector for packaging into a virus (e.g. retroviruses, replication defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated viruses). Viral vectors also include polynucleotides carried by a virus for transfection into a host cell.

Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g. bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors." Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

Recombinant expression vectors can comprise a nucleic acid of the presently disclosed subject matter in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory elements, which may be selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed.

Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence (e.g. in an in vitro transcription/translation system or in a host cell when the vector is introduced into the host cell).

The term "regulatory element" is intended to include promoters, enhancers, internal ribosomal entry sites (IRES), and other expression control elements (e.g. transcription termination signals, such as polyadenylation signals and poly-U sequences). Such regulatory elements are described, for example, in Goeddel (1990) Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. Regulatory elements include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain

host cells (e.g., tissue-specific regulatory sequences). A tissue-specific promoter may direct expression primarily in a desired tissue of interest, such as muscle, neuron, bone, skin, blood, specific organs (e.g. liver, pancreas), or particular cell types (e.g. lymphocytes). Regulatory elements may also direct expression in a temporal-dependent manner, such as in a cell-cycle dependent or developmental stage-dependent manner, which may or may not also be tissue or cell-type specific.

In some embodiments, a vector comprises one or more pol III promoters, one or more pol II promoters, one or more pol I promoters, or combinations thereof. Examples of pol III promoters include, but are not limited to, U6 and H1 promoters. Examples of pol II promoters include, but are not limited to, the retroviral Rous sarcoma virus (RSV) LTR promoter (optionally with the RSV enhancer), the cytomegalovirus (CMV) promoter (optionally with the CMV enhancer) (e.g., Boshart et al. (1985) *Cell* 41:521-530), the SV40 promoter, the dihydrofolate reductase promoter, the β-actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1α promoter.

Also encompassed by the term "regulatory element" are enhancer elements, such as WPRE; CMV enhancers; the R-U5′ segment in LTR of HTLV-I (Takebe et al. (1988) *Mol. Cell. Biol.*8:466-472); SV40 enhancer; and the intron sequence between exons 2 and 3 of rabbit β-globin (O'Hare et al. (1981) *Proc. Natl. Acad. Sci. USA.* 78(3):1527-31). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression desired, etc. A vector can be introduced into host cells to thereby produce transcripts, proteins, or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., clustered regularly interspersed short palindromic repeats (CRISPR) transcripts, proteins, enzymes, mutant forms thereof, fusion proteins thereof, etc.). Advantageous vectors include lentiviruses and adeno-associated viruses, and types of such vectors can also be selected for targeting particular types of cells.

The terms "polynucleotide", "nucleotide", "nucleotide sequence", "nucleic acid" and "oligonucleotide" are used interchangeably. They refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Polynucleotides may have any three dimensional structure, and may perform any function, known or unknown. The following are non-limiting examples of polynucleotides: coding or non-coding regions of a gene or gene fragment, loci (locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, short

interfering RNA (siRNA), short-hairpin RNA (shRNA), micro-RNA (miRNA), ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. A polynucleotide may comprise one or more modified nucleotides, such as methylated nucleotides and nucleotide analogs. If present, modifications to the nucleotide structure may be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component.

In aspects of the presently disclosed subject matter the terms "chimeric RNA", "chimeric guide RNA", "guide RNA", "single guide RNA", "synthetic guide RNA" and "crRNA" are used interchangeably and refer to the polynucleotide sequence comprising the guide sequence. The term "guide sequence" refers to the about 20 bp sequence within the guide RNA that specifies the target site and may be used interchangeably with the terms "guide" or "spacer".

As used herein the term "wild type" is a term of the art understood by skilled persons and means the typical form of an organism, strain, gene or characteristic as it occurs in nature as distinguished from mutant or variant forms.

As used herein the term "variant" should be taken to mean the exhibition of qualities that have a pattern that deviates from what occurs in nature.

The terms "non-naturally occurring" or "engineered" are used interchangeably and indicate the involvement of the hand of man. The terms, when referring to nucleic acid molecules or polypeptides mean that the nucleic acid molecule or the polypeptide is at least substantially free from at least one other component with which they are naturally associated in nature and as found in nature.

"Complementarity" refers to the ability of a nucleic acid to form hydrogen bond(s) with another nucleic acid sequence by either traditional Watson-Crick or other non-traditional types. A percent complementarity indicates the percentage of residues in a nucleic acid molecule which can form hydrogen bonds (e.g., Watson-Crick base pairing) with a second nucleic acid sequence (e.g., 5, 6, 7, 8, 9, 10 out of 10 being 50%, 60%, 70%, 80%, 90%, and 100% complementary). "Perfectly complementary" means that all the contiguous residues of a nucleic acid sequence will hydrogen bond with the same number of contiguous residues in a second nucleic acid sequence. "Substantially complementary" as used herein refers to a degree of complementarity that is at least 60%, 65%, 70%, 75%,

80%, 85%, 90%, 95%. 97%, 98%, 99%, or 100% over a region of 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, or more nucleotides, or refers to two nucleic acids that hybridize under stringent conditions.

As used herein, "stringent conditions" for hybridization refer to conditions under which a nucleic acid having complementarity to a target sequence predominantly hybridizes with the target sequence, and substantially does not hybridize to non-target sequences. Stringent conditions are generally sequence-dependent, and vary depending on a number of factors. In general, the longer the sequence, the higher the temperature at which the sequence specifically hybridizes to its target sequence. Non-limiting examples of stringent conditions are described in detail in Tijssen (1993), Laboratory Techniques In Biochemistry And Molecular Biology-Hybridization With Nucleic Acid Probes Part 1, Second Chapter "Overview of principles of hybridization and the strategy of nucleic acid probe assay", Elsevier, N.Y.

"Hybridization" refers to a reaction in which one or more polynucleotides react to form a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues. The hydrogen bonding may occur by Watson Crick base pairing, Hoogstein binding, or in any other sequence specific manner. The complex may comprise two strands forming a duplex structure, three or more strands forming a multi stranded complex, a single self hybridizing strand, or any combination of these. A hybridization reaction may constitute a step in a more extensive process, such as the initiation of PCR, or the cleavage of a polynucleotide by an enzyme. A sequence capable of hybridizing with a given sequence is referred to as the "complement" of the given sequence.

As used herein, "expression" refers to the process by which a polynucleotide is transcribed from a DNA template (such as into and mRNA or other RNA transcript) and/or the process by which a transcribed mRNA is subsequently translated into peptides, polypeptides, or proteins. Transcripts and encoded polypeptides may be collectively referred to as "gene product." If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in a eukaryotic cell.

The terms "polypeptide", "peptide" and "protein" are used interchangeably herein to refer to polymers of amino acids of any length. The polymer may be linear or branched, it may comprise modified amino acids, and it may be interrupted by non amino acids. The terms also encompass an amino acid polymer that has been modified; for example, disulfide

bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation, such as conjugation with a labeling component.

As used herein the term "amino acid" includes natural and/or unnatural or synthetic amino acids, including glycine and both the D or L optical isomers, and amino acid analogs and peptidomimetics.

The practice of the present presently disclosed subject matter employs, unless otherwise indicated, conventional techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics and recombinant DNA, which are within the skill of the art (Sambrook, Fritsch and Maniatis (1989) Molecular Cloning: A Laboratory Manual, 2nd edition; Ausubel et al., eds. (1987) Current Protocols in Molecular Biology); MacPherson et al., eds. (1995) Methods in Enzymology (Academic Press, Inc.): PCR 2: A Practical Approach); Harlow and Lane, eds. (1988) Antibodies, A Laboratory Manual; Freshney, ed. (1987) Animal Cell Culture).

Several aspects of the presently disclosed subject matter relate to vector systems comprising one or more vectors, or vectors as such. Vectors can be designed for expression of CRISPR transcripts (e.g. nucleic acid transcripts, proteins, or enzymes) in prokaryotic or eukaryotic cells. For example, CRISPR transcripts can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors), yeast cells, or mammalian cells. Suitable host cells are discussed further in Goeddel (1990) Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Vectors may be introduced and propagated in a prokaryote. In some embodiments, a prokaryote is used to amplify copies of a vector to be introduced into a eukaryotic cell or as an intermediate vector in the production of a vector to be introduced into a eukaryotic cell (e.g. amplifying a plasmid as part of a viral vector packaging system). In some embodiments, a prokaryote is used to amplify copies of a vector and express one or more nucleic acids, such as to provide a source of one or more proteins for delivery to a host cell or host organism. Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins.

Fusion vectors add a number of amino acids to a protein encoded therein, such as to the amino terminus of the recombinant protein. Such fusion vectors may serve one or more

purposes, such as: (i) to increase expression of recombinant protein; (ii) to increase the solubility of the recombinant protein; and (iii) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Example fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson (1988) *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A. respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann et al. (1988) *Gene* 69:301-315) and pET 11d (Studier et al. (1990) Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif.).

In some embodiments, a vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerivisae* include pYepSec1 (Baldari, et al.(1987) *EMBO J.* 6: 229-234), pMFa (Kuijan and Herskowitz (1982) *Cell* 30: 933-943), pJRY88 (Schultz et al. (1987) *Gene* 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

In some embodiments, a vector is capable of driving expression of one or more sequences in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed (1987) *Nature* 329: 840) and pMT2PC (Kaufman et al. (1987) *EMBO J.* 6: 187-195). When used in mammalian cells, the expression vector's control functions are typically provided by one or more regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, simian virus 40, and others disclosed herein and known in the art. For other suitable expression systems for both prokaryotic and eukaryotic cells see, e.g., Chapters 16 and 17 of Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y..

In some embodiments, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific

regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al. (1987) *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.*8: 729-733) and immunoglobulins (Baneiji et al. (1983) *Cell* 33: 729-740; Queen and Baltimore (1983) *Cell* 33: 741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (Edlund et al.(1985) *Science* 230: 912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, e.g., the murine hox promoters (Kessel and Gruss (1990) *Science* 249: 374-379) and the α-fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3: 537-546).

In some embodiments, a regulatory element is operably linked to one or more elements of a CRISPR system so as to drive expression of the one or more elements of the CRISPR system. In general, CRISPRs (Clustered Regularly Interspaced Short Palindromic Repeats), also known as SPIDRs (SPacer Interspersed Direct Repeats), constitute a family of DNA loci that are usually specific to a particular bacterial species. The CRISPR locus comprises a distinct class of interspersed short sequence repeats (SSRs) that were recognized in *E. coli* (Ishino et al. (1987) *J. Bacteriol.*, 169:5429-5433; and Nakata et al. (1989) J. *Bacteriol.*, 171:3553-3556), and associated genes. Similar interspersed SSRs have been identified in *Haloferax mediterranei, Streptococcus pyogenes, Anabaena,* and *Mycobacterium tuberculosis* (Groenen et al. (1993) *Mol. Microbiol.*, 10:1057-1065; Hoe et al. (1999) *Emerg. Infect. Dis.*, 5:254-263; Masepohl et al. (1996) *Biochim. Biophys. Acta* 1307:26-30; and Mojica et al. (1995) *Mol. Microbiol.*, 17:85-93). The CRISPR loci typically differ from other SSRs by the structure of the repeats, which have been termed short regularly spaced repeats (SRSRs) (Janssen et al. (2002) *OMICS J. Integ. Biol.*, 6:23-33; and Mojica et al. (2000) *Mol. Microbiol.*, 36:244-246). In general, the repeats are short elements that occur in clusters that are regularly spaced by unique intervening sequences with a substantially constant length (Mojica et al. (2000) *Mol. Microbiol.*, 36:244-246). Although the repeat sequences are highly conserved between strains, the number of interspersed repeats and the sequences of the spacer regions typically differ from strain to strain (van Embden et al. (2000) *J. Bacteriol.*, 182:2393-2401). CRISPR loci have been

identified in more than 40 prokaryotes (e.g., Jansen et al. (2002) *Mol. Microbiol.*, 43:1565-1575; and Mojica et al. (2005) *J. Mol. Evol.* 60:174-82) including, but not limited to *Aeropyrum, Pyrobaculum, Sulfolobus, Archaeoglobus, Halocarcula, Methanobacteriumn, Methanococcus, Methanosarcina, Methanopyrus, Pyrococcus, Picrophilus, Thernioplasnia, Corynebacterium, Mycobacterium, Streptomyces, Aquifrx, Porphvromonas, Chlorobium, Thermus, Bacillus, Listeria, Staphylococcus, Clostridium, Thermoanaerobacter, Mycoplasma, Fusobacterium, Azarcus, Chromobacterium, Neisseria, Nitrosomonas, Desulfovibrio, Geobacter, Myrococcus, Campylobacter, Wolinella, Acinetobacter, Erwinia, Escherichia, Legionella, Methylococcus, Pasteurella, Photobacterium, Salmonella, Xanthomonas, Yersinia, Treponema*, and *Thermotoga.*

In general, "CRISPR system" refers collectively to transcripts and other elements involved in the expression of or directing the activity of CRISPR-associated ("Cas") genes, including sequences encoding a Cas gene, a guide sequence (also referred to as a "spacer" in the context of an endogenous CRISPR system), or other sequences and transcripts from a CRISPR locus. In some embodiments, one or more elements of a CRISPR system is derived from a type I, type II, or type III CRISPR system. In some embodiments, one or more elements of a CRISPR system is derived from a particular organism comprising an endogenous CRISPR system, such as *Streptococcus pyogenes*. In general, a CRISPR system is characterized by elements that promote the formation of a CRISPR complex at the site of a target sequence (also referred to as a protospacer in the context of an endogenous CRISPR system).

In the context of formation of a CRISPR complex, "target sequence" refers to a sequence to which a guide sequence is designed to have complementarity, where hybridization between a target sequence and a guide sequence promotes the formation of a CRISPR complex. Full complementarity is not necessarily required, provided there is sufficient complementarity to cause hybridization and promote formation of a CRISPR complex. A target sequence may comprise any polynucleotide, such as DNA or RNA polynucleotides. In some embodiments, a target sequence is located in the nucleus or cytoplasm of a cell. In some embodiments, the target sequence may be within an organelle of a eukaryotic cell, for example, mitochondrion or chloroplast. A sequence or template that may be used for recombination into the targeted locus comprising the target sequences is referred to as an "editing template" or "editing polynucleotide" or "editing sequence". In aspects of the presently disclosed subject matter, an exogenous template polynucleotide

may be referred to as an editing template. In an aspect of the presently disclosed subject matter the recombination is homologous recombination.

In some embodiments, a vector comprises one or more insertion sites, such as a restriction endonuclease recognition sequence (also referred to as a "cloning site"). In some embodiments, one or more insertion sites (e.g. about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more insertion sites) are located upstream and/or downstream of one or more sequence elements of one or more vectors. When multiple different guide sequences are used, a single expression construct may be used to target CRISPR activity to multiple different, corresponding target sequences within a cell. For example, a single vector may comprise about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, or more guide sequences. In some embodiments, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more such guide-sequence-containing vectors may be provided, and optionally delivered to a cell.

In some embodiments, a vector comprises a regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme, such as a Cas protein. Non-limiting examples of Cas proteins include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 and Csx12), Cas10, Cas13a, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, homologs thereof, or modified versions thereof. These enzymes are known; for example, the amino acid sequence of *S. pyogenes* Cas9 protein may be found in the SwissProt database under accession number Q99ZW2. In some embodiments, the unmodified CRISPR enzyme has DNA cleavage activity, such as Cas9. In some embodiments the CRISPR enzyme is Cas9, and may be Cas9 from *S. pyogenes* or *S. pneumoniae*.

In some embodiments, the CRISPR enzyme directs cleavage of one or both strands at the location of a target sequence, such as within the target sequence and/or within the complement of the target sequence. In some embodiments, the CRISPR enzyme directs cleavage of one or both strands within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, or more base pairs from the first or last nucleotide of a target sequence. In some embodiments, a vector encodes a CRISPR enzyme that is mutated to with respect to a corresponding wild-type enzyme such that the mutated CRISPR enzyme lacks the ability to cleave one or both strands of a target polynucleotide containing a target sequence.

In some embodiments, an enzyme coding sequence encoding a CRISPR enzyme is codon optimized for expression in particular cells, such as eukaryotic cells. The eukaryotic cells may be those of or derived from a particular organism, such as a mammal, including but not limited to human, mouse, rat, rabbit, dog, or non-human primate. 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", and these tables can be adapted in a number of ways. See Nakamura et al. (2000) *Nucl. Acids Res.* 28:292. 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 CRISPR enzyme correspond to the most frequently used codon for a particular amino acid.

In general, a guide sequence is any polynucleotide sequence having sufficient complementarity with a target polynucleotide sequence to hybridize with the target sequence and direct sequence-specific binding of a CRISPR complex to the target sequence. In some embodiments, the degree of complementarity between a guide sequence and its corresponding target sequence, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g. the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT,

Novoalign (Novocraft Technologies, ELAND (Illumina, San Diego, Calif.), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). In some embodiments, a guide sequence is about or more than 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, or more nucleotides in length. In some embodiments, a guide sequence is less than about 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, or fewer nucleotides in length.

The ability of a guide sequence to direct sequence-specific binding of a CRISPR complex to a target sequence may be assessed by any suitable assay. For example, the components of a CRISPR system sufficient to form a CRISPR complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target sequence, such as by transfection with vectors encoding the components of the CRISPR sequence, followed by an assessment of preferential cleavage within the target sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target polynucleotide sequence may be evaluated in a test tube by providing the target sequence, components of a CRISPR complex, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art.

A guide sequence may be selected to target any target sequence. In some embodiments, the target sequence is a sequence within a genome of a cell. Exemplary target sequences include those that are unique in the target genome.

In some embodiments, the CRISPR enzyme is part of a fusion protein comprising one or more heterologous protein domains (e.g. about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more domains in addition to the CRISPR enzyme). A CRISPR enzyme fusion protein may comprise any additional protein sequence, and optionally a linker sequence between any two domains. Examples of protein domains that may be fused to a CRISPR enzyme include, without limitation, epitope tags, reporter gene sequences, and protein domains having one or more of the following activities: methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity and nucleic acid binding activity. Non-limiting examples of epitope tags include histidine (His) tags, V5 tags, FLAG tags, influenza hemagglutinin (HA) tags, Myc tags, VSV-G tags, and thioredoxin (Trx) tags. Examples of reporter genes include, but are not limited to,

glutathione-5-transferase (GST), horseradish peroxidase (HRP), chloramphenicol acetyltransferase (CAT) beta-galactosidase, beta-glucuronidase, luciferase, green fluorescent protein (GFP), HcRed, DsRed, cyan fluorescent protein (CFP), yellow fluorescent protein (YFP), and autofluorescent proteins including blue fluorescent protein (BFP). A CRISPR enzyme may be fused to a gene sequence encoding a protein or a fragment of a protein that bind DNA molecules or bind other cellular molecules, including but not limited to maltose binding protein (MBP), S-tag, Lex A DNA binding domain (DBD) fusions, GAL4A DNA binding domain fusions, and herpes simplex virus (HSV) BP16 protein fusions. Additional domains that may form part of a fusion protein comprising a CR ISPR enzyme are described in US20110059502, incorporated herein by reference. In some embodiments, a tagged CRISPR enzyme is used to identify the location of a target sequence.

In an aspect of the presently disclosed subject matter, a reporter gene which includes but is not limited to glutathione-5-transferase (GST), horseradish peroxidase (HRP), chloramphenicol acetyltransferase (CAT) beta-galactosidase, beta-glucuronidase, luciferase, green fluorescent protein (GFP), HcRed, DsRed, cyan fluorescent protein (CFP), yellow fluorescent protein (YFP), and autofluorescent proteins including blue fluorescent protein (BFP), may be introduced into a cell to encode a gene product which serves as a marker by which to measure the alteration or modification of expression of the gene product. In a further embodiment of the presently disclosed subject matter, the DNA molecule encoding the gene product may be introduced into the cell via a vector. In a preferred embodiment of the presently disclosed subject matter the gene product is luciferase. In a further embodiment of the presently disclosed subject matter the expression of the gene product is decreased.

Generally, promoter embodiments of the present presently disclosed subject matter comprise: 1) a complete Pol III promoter, which includes a TATA box, a Proximal Sequence Element (PSE), and a Distal Sequence Element (DSE); and 2) a second basic Pol III promoter that includes a PSE and TATA box fused to the 5′ terminus of the DSE in reverse orientation. The TATA box, which is named for its nucleotide sequence, is a major determinant of Pol III specificity. It is usually located at a position between nt. −23 and −30 relative to the transcribed sequence, and is a primary determinant of the beginning of the transcribed sequence. The PSE is usually located between nt. −45 and −66. The DSE

enhances the activity of the basic Pol III promoter. In the H1 promoter, there is no gap between the PSE and the DSE.

Bidirectional promoters refer to any promoter (typically pol II) that are orientated in head-tail-tail-head fasion –basically any control region that direct divergent transcription. In some embodiments, the bidirectional H1 promoter could be pol II and pol II or pol III and pol III, or combinations thereof. The H1 bidirectional, as well as 7Sk and others described herein, are unique in that they are bidirectional, which is itself uncommon in eukaryotic genomes, but most importantly here, directs a protein coding gene on one side (pol II), and an RNA gene on the other side (pol III).

In some embodiments, bidirectional promoters consists of: 1) a complete, conventional, unidirectional Pol III promoter that contains 3 external control elements: a DSE, a PSE, and a TATA box; and 2) a second basic Pol III promoter that includes a PSE and a TATA box fused to the 5′ terminus of the DSE in reverse orientation. The TATA box, which is recognized by the TATA binding protein, is essential for recruiting Pol III to the promoter region. Binding of the TATA binding protein to the TATA box is stabilized by the interaction of SNAPc with the PSE. Together, these elements position Pol III correctly so that it can transcribe the expressed sequence. The DSE is also essential for full activity of the Pol III promoter (Murphy et al. (1992) *Mol. Cell Biol.* 12:3247-3261; Mittal et al. (1996) *Mol. Cell Biol.* 16:1955-1965; Ford and Hernandez (1997) *J.Biol.Chem.*, 272:16048-16055; Ford et al. (1998) *Genes, Dev.*, 12:3528-3540; Hovde et al. (2002) *Genes Dev.* 16:2772-2777). Transcription is enhanced up to 100-fold by interaction of the transcription factors Oct-1 and/or SBF/Staf with their motifs within the DSE (Kunkel and Hixon (1998) *Nucl. Acid Res.*, 26:1536-1543). Since the forward and reverse oriented basic promoters direct transcription of sequences on opposing strands of the double-stranded DNA templates, the positive strand of the reverse oriented basic promoter is appended to the 5′ end of the negative strand of the DSE. Transcripts expressed under the control of the H1 promoter are terminated by an unbroken sequence of 4 or 5 T's.

In the H1 promoter, the DSE is adjacent to the PSE and the TATA box (Myslinski et al. (2001) *Nucl. Acid Res.* 29:2502-2509). To minimize sequence repetition, this promoter was rendered bidirectional by creating a hybrid promoter, in which transcription in the reverse direction is controlled by appending a PSE and TATA box derived from the U6 promoter. To facilitate construction of the bidirectional H1 promoter, a

small spacer sequence may also inserted between the reverse oriented basic promoter and the DSE.

Examples of cell cycle dependent proteins may comprise any of the following:

(a) Chromosomal DNA Replication Reaction Including Initiation of Chromosomal DNA Replication and Progression of Replication Fork

Mcm10, Orc1, Orc3, Cdc6, Cdt1, Geminin, Mcm3, Mcm4, Mcm5, Mcm6, Mcm7, Mcm8, Cdc7, Cdc5, Psf1, Psf2, Psf3, Cdc45, Pola p180, Pola p70, Pola Spp1(Prim2a), RPA70, RPA34, PCNA, Elg1, Ligase1, Pole Pol2, Pole Dpb3, Topoisomerase I, TDP1, Orc2, Orc4, Orc5, Orc6, Mcm2, Dbf4, TopBP1, Sld5, Pola Spp2, RFC1, RFC2, RFC3, RFC4, RFC5, Pif1, Pold p50, Pole Dpb2, Topoisomerase Iia, Topoisomerase Iib, RPAI4, FEN1, DNA2, Pold p125, Pold p68, Pold p12, Pole Dpb4

(b) DNA Damage Checkpoints

ATR, Chk1, NBS1, Hus1, Rad1, Mad2, BubR1, ATM, Rad50, Mre11, Mdc1, 53BP1, Rad17, BubR1, ATRIP, Chk2, H2AX, RFC1, RFC2, RFC3, RFC4, RFC5, ATM, BRCA1, Chk1, Chk2, 14-3-3eta, 14-3-3sigma, cdc25A, cdc25c, wee1, ATR, ATRIP, Rad17, RFC2, RFC3, RFC4, RFC5, HUS1, Rad1, Rad9, P53, Rad50, Mre11, NBS1, TopBP1, 53BP1, H2AX

(c) Sister Chromatid Agglutination and Separation

Ctf18, Scc1, Scc3, Dcc1, Trf4-1, Trf4-2, Smc1, Smc3, Pds1(Securin), Mad-2, BubR1, Esp1

(d) Base Excision Repair

UNG, MBD4, TDG, NTH1, NEIL2, NEIL3, APE2, PARP1, PNK, Polb, OGG1, APE1, XRCC1, Ligase3, SMUG1, TDG, MYH, MPG, NEIL1, ADPRT, ADPRTL2, MGMT, ABH1, ABH2, ABH3

(e) Mismatch Excision Repair

MSH2, PMS1, PMS2, MLH3, Exonuclease1, MSH3, MSH6, MSH5, MLH1, MSH4, PMS2L3, Trex1, Trex2, PMS2L4

(f) Nucleotide Excision Repair

XPC, Rad23A, Rad23B, CSA, CSB, XPG, XPF, DDB1, DDB2, XAB2, XPB, ERCC1, XPD, XPA, DDB2, Mms19, CETN2, RPA70, RPA34, RPAI4, GTF2H1, GTF2H2, GTF2H3, GTF2H4, CDK7, CCNH, MNAT1, LigaseI, CSA, CSB

(g) Homologous Recombination Repair

Rad51, Rad51L1, Rad51C, Rad51L3, DMC1, XRCC2, XRCC3, Rad52, Rad54L, Rad54B, BRCA1, BRCA2, Rad50, Mre11, NBS1, TopoisomeraseIIIa, TopoisomeraseIIIb, WHIP, WRN, BLM, RecQ1, RecQ5

(h) Non-Homologous End-Joining Repair (Non-Homologous Recombination Repair)

Ku70, Ku80, DNA-pk, Ligase4, XRCC4, Artemis, WRN

i) Double-Strand DNA Break Repair

Rad51, Rad51D, Xrcc2, Rad54, BRCA1, Ku80, XRCC4, Rad52, Rad51C, Dmc1, Rad54B, DNA-pk, Ku70, Ligase4, Rad51B, XRCC3, BRCA2, Artemis

(j) DNA Post-Replication Repair (DNA Damage Tolerance)

Rad6A, Rad6B, Rad18, Ubc13, FBH1

(k) DNA Crosslink Damage Repair

FANCA, FANCC, FANCD2, FANCE, FANCF, FANCG

(l) DNA-Protein Crosslink Damage Repair

TDP1

(m) DNA Polymerase

Poli, Polh, Polq, Polk, Polz (REV3), Poll, Polm, Rev1, Polb, Polg, Pold p50, Pole Pol2, REV7, Poln, Pola P180, Pola p70, Pola Spp1, Pola Spp2, Pold p68, Pold p12, Pole Dpb2, Pole Dpb3, Pole Dpb4

(n) Nucleotide Cleansing

MTH1, DUT, p53R2

(o) Chromatin Structure Maintenance

H2AX, Sir2, SIRT1 (Sirtuin)

(p) Telomere Structure Maintenance

Tin2, Sir2, hTert, TRF1, TRF2, Tankyrase, Pot1, Rap1, Pif1

Preferred examples of genes associated with each of the aforementioned functions (a) to (p) include the genes described in Examples below. More specifically, examples of such genes are as follows: APE2, ATR, BRCA1, Chk1, Cdc5, Cdc6, Cdc7, Cdc45, Cdt1, CSA, CSB, Ctf18, DDB1, DDB2, DNA2, DUT, Elg1, EndoV, Esp1, Exonuclease1, FBH1, FEN1, Geminin, Hus1, KNTC2 (NDC80), Ku80, Ligase1, Mad2, MBD4, Mcm3, Mcm4, Mcm5, Mcm6, Mcm7, Mcm8, Mcm10, MGMT, MLH3, Mms4, MPG, MSH2, Mus81, NBS1, NEIL2, NEIL3, NTH1, Orc1, Orc3, PARP1, PCNA, Pif1, PMS1, PMS2, PNK, Pola p180, Pola p70, Pola Spp1(Prim2a), Polb, Pold p125, Pole Dpb3, Pole Dpb4, Pole Pol2, Poli, Poll, Polm, Psf1, Psf2, Psf3, Rad1, Rad18, Rad23A, Rad23B, Rad51, Rad51D, Rad54,

Rad6A, RPA34, RPA70, Scc1, Scc3, Sir2, SIRT1 (Sirtuin), TDG, TDP1, TIMELESS, Tin2, Topoisomerase I, Topoisomerase IIIa, Topoisomerase IIIb, Ubc13, UNG, XAB2, XPC, XPF, XPG, Xrcc2, and XRCC4.

Since the gene names described in the present specification are names which are widely and generally known, those skilled in the art are able to suitably acquire data on the nucleotide sequences of said genes from a public reference database or gene database (e.g., GenBank) based on the gene name.

B. Methods

In some embodiments, the presently disclosed subject matter also provides a method of altering expression of one or more gene products in a eukaryotic or non-eukaryotic cell, wherein the cell comprises a DNA molecule encoding the one or more gene products, the method comprising introducing into the cell an improved and modified non-naturally occurring CRISPR-Cas system previously described in WO2015/195621 (herein incorporated by reference in its entirety). Such improvements comprise a non-naturally occurring nuclease system (e.g., CRISPR-Cas9) comprising one or more vectors comprising: a) a promoter (e.g., orthogolous H1 promoter or 7sk) operably linked to at least one nucleotide sequence encoding a nuclease system guide RNA (gRNA), wherein the gRNA hybridizes with a target sequence of a DNA molecule in a cell, and wherein the DNA molecule encodes one or more gene products expressed in the cell; and b) a regulatory element operable in a cell operably linked to a nucleotide sequence encoding a genome-targeted nuclease (e.g., modified Cas9 protein), wherein components (a) and (b) are located on the same or different vectors of the system, wherein the gRNA targets and hybridizes with the target sequence and the nuclease cleaves the DNA molecule to alter expression of the one or more gene products. In some embodiments, the system is packaged into a single adeno-associated virus (AAV) particle. In some embodiments, the system inactivates one or more gene products. In some embodiments, the system excises at least one gene mutation. In some embodiments, the H1 promoter comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 12 or 32. In some embodiments, the orthologous H1 promoter is derived from mouse or rat. In some embodiments, the orthologous H1 promoter comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NOs: 84-119. In some embodiments, the orthologous H1 promoter comprises a nucleotide sequences

set forth in the group consisting of SEQ ID NOs: 84-119. In some embodiments, the H1 promoter comprises: a) control elements that provide for transcription in one direction of at least one nucleotide sequence encoding a gRNA; and b) control elements that provide for transcription in the opposite direction of a nucleotide sequence encoding a genome-targeted nuclease. In some embodiments, the promoter is the human, mouse, or rat 7sk promoter. In some embodiments, the 7sk is selected from the group consisting of 7sk1 (variant 1 of 7sk which uses the endogenous GSTA4 5'UTR), 7sk2 (variant 2 of 7sk which uses just a kozak sequence in the 5'UTR), and 7sk3 (variant of 7sk which uses the beta globin 5'UTR). In some embodiments, the human 7sk1 comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 3. In some embodiments, the mouse 7sk1 comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to SEQ ID NO: 6. In some embodiments, the human 7sk2 comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to SEQ ID NO: 4. In some embodiments, the mouse 7sk2 comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to SEQ ID NO: 7. In some embodiments, the human 7sk3 comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to SEQ ID NO: 5. In some embodiments, the mouse 7sk3 comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to SEQ ID NO: 8. In some embodiments, the orthologous H1 promoter or 7sk promoter has at least one modification. In some embodiments, the at least one modification of the promoter comprises an element that allows conditional regulation. In some embodiments, the element is a tet-responsive promoter. In some embodiments, the tet-response promoter comprises a Tet repressor (TetR) and Tet operator (TetO) engineered into the H1 promoter. In some embodiments, the TetR comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to SEQ ID NO: 14. In some embodiments, the TetO comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 11. In some embodiments, the H1-TetO comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 13. In some embodiments, the at least one modification of the promoter comprises a site that allows auto-regulation. In some embodiments, the auto-regulation site comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to

the nucleotide sequence set forth in SEQ ID NO: 33 or 34. In some embodiments, the nuclease system further comprising at least one terminator sequence. In some embodiments, the at least one terminator sequence is selected from the group consisting of SV40 or synthetic poly A (SPA) sequences. In some embodiments, the terminator sequences is a SV40 120 (SEQ ID NO: 123) or 240 (SEQ ID NO: 122) base pair sequence. In some embodiments, the terminator sequences is a SPA 49 base pair equence. In some embodiments, the nuclease system further comprises 5' untranslated region (5'UTR) sequences. In some embodiments, the nuclease system further comprises a Kozak sequence. In some embodiments, the Kozak sequence comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to SEQ ID NO: 1 or 2. In some embodiments, the nuclease system further comprises a RNA sequence that mediates cap-independent initiation of translation. In some embodiments, the RNA sequence is selected from the group consisting of 6.947 or 6.967. In some embodiments, the Cas9 comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 61. In some embodiments, the Cas9 comprises at least one modification. In some embodiments, the at least one modification in the Cas9 comprises an alteration in the cleaving sequence. In some embodiments, the nuclease is seleted from the group consisting of Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 and Csx12), Cas10, Cas13a, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, Cpf1, C2c1, C2c2, and C2c3. In some embodiments, the alteration in the cleaving sequence selected from the group consisting of T2A, P2A, E2A, and F2A. In some embodiments, the T2A comprises an amino acid sequence. having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 36. In some embodiments, the P2A comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 35. In some embodiments, the P2A comprises an amino acid sequence having the nucleotide sequence set forth in SEQ ID NO: 35. In some embodiments, the E2A comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 37. In some embodiments, the F2A comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or

100% identity to the nucleotide sequence set forth in SEQ ID NO: 38. In some embodiments, the at least one modification in the Cas9 comprises a codon optimized for expression in the cell. In some embodiments, the Cas9 further comprises a linker sequence operably fused in frame to a cell-cycle dependent protein (Cas9 fusion). In some embodiments, the cell cycle-dependent protein is selected from the group consisting of APE2, ATR, BRCA1, Chk1, Cdc5, Cdc6, Cdc7, Cdc45, Cdt1, CSA, CSB, Ctf18, DDB1, DDB2, DNA2, DUT, Elg1, EndoV, Esp1, Exonuclease1, FBH1, FEN1, Geminin, Hus1, KNTC2 (NDC80), Ku80, Ligase1, Mad2, MBD4, Mcm3, Mcm4, Mcm5, Mcm6, Mcm7, Mcm8, Mcm10, MGMT, MLH3, Mms4, MPG, MSH2, Mus81, NBS1, NEIL2, NEIL3, NTH1, Orc1, Orc3, PARP1, PCNA, Pif1, PMS1, PMS2, PNK, Pola p180, Pola p70, Pola Spp1 (Prim2a), Polb, Pold p125, Pole Dpb3, Pole Dpb4, Pole Pol2, Poli, Poll, Polm, Psf1, Psf2, Psf3, Rad1, Rad18, Rad23A, Rad23B, Rad51, Rad51D, Rad54, Rad6A, RPA34, RPA70, Scc1, Scc3, Sir2, SIRT1 (Sirtuin), TDG, TDP1, TIMELESS, Tin2, Topoisomerase I, Topoisomerase IIIa, Topoisomerase IIIb, Ubc13, UNG, XAB2, XPC, XPF, XPG, Xrcc2, and XRCC4. In some embodiments, the cell cycle-dependent protein is Geminin. In some embodiments, the cell cycle-dependent protein is human Geminin. In some embodiments, the human Geminin comprises the amino acids from positions 1-110 (hGem(1-110)). In some embodiments, the hGem(1-110) comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the amimo acid sequence set forth in SEQ ID NO: 19. In some embodiments, the cell cycle-dependent protein is Cdt1. In some embodiments, the cell cycle-dependent protein is human Cdt1. In some embodiments, the human Cdt1 comprises amino acids from positions 30-120 (hCdt1(30-120)). In some embodiments, the hCdt1(30-120) comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the amino acid sequence set forth in SEQ ID NO: 18. In some embodiments, the Cas9 fusion comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the amino acid sequence selected from the group consisting of SEQ ID NO: 20-27. In some embodiments, the Cas9 is operably fused in frame to a ubiquitin protein (Ub-Cas9). In some embodiments, the Ub-Cas9 at least one N-terminal modification. In some embodiments, the N-terminal modified Ub-Cas9 comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the amino acid sequence selected from the group consisting of SEQ ID NO: 39-58. In some embodiments, the ubiquitin protein comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100%

identity to the nucleotide sequence set forth in SEQ ID NO: 60. In some embodiments, the nuclease system further comprises a SaCas9 nickase. In some embodiments, the nuclease system further comprises a donor template sequence. In some embodiments, the at least one vector comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 62 or 67. In some embodiments, the donor template sequence corrects at least one gene mutation. In some embodiments, the at least one gene mutation is *rd10* or *rd12*. In some embodiments, the at least one vector comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 63 or 64. In some embodiments, the at least one vector comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 65 or 66. In some embodiments, the promoter is operably linked to at least one, two, three, four, five, six, seven, eight, nine, or ten gRNA. In some embodiments, the target sequence comprises the nucleotide sequence $AN_{19}NGG$, $GN_{19}NGG$, $CN_{19}NGG$, or $TN_{19}NGG$. In some embodiments, the cell is a eukaryotic or or non-eukaryotic cell. In some embodiments, the eukaryotic cell is a mammalian or human cell. In some embodiments, the eukaryotic cell is a retinal photoreceptor cell. In some embodiments, the one or more gene products are rhodopsin. In some embodiments, the expression of the one or more gene products is decreased. In some embodiments, the system inactivates one or more gene products. In some embodiments, the nuclease system excises at least one gene mutation. In some embodiments, the expression of the one or more gene products is decreased. In some embodiments, the cell is a retinal ganglion cell. In some embodiments, the eukaryotic cell is a cancerous cell. In some embodiments, cell proliferation is inhibited or reduced in the cancerous cell. In some embodiments, the apoptosis is enhanced or increased in the cancerous cell.

In some embodiments, the presently disclosed subject matter also provides a method of altering expression of one or more gene products in a cell, wherein the cell comprises a DNA molecule encoding the one or more gene products, the method comprising introducing into the cell a non-naturally occurring CRISPR-Cas system comprising one or more vectors comprising: a) an H1 promoter operably linked to at least one nucleotide sequence encoding a CRISPR-Cas system guide RNA (gRNA), wherein the gRNA hybridizes with a target sequence of the DNA molecule; and b) a regulatory element operable in the cell operably linked to a nucleotide sequence encoding a Cas9 protein,

wherein components (a) and (b) are located on the same or different vectors of the system, wherein the gRNA targets and hybridizes with the target sequence and the Cas9 protein cleaves the DNA molecule to alter expression of the one or more gene products.

In some embodiments, the presently disclosed subject matter also provides a method of altering expression of one or more gene products in a eukaryotic cell, wherein the cell comprises a DNA molecule encoding the one or more gene products, the method comprising introducing into the cell a non-naturally occurring CRISPR-Cas system comprising one or more vectors comprising: a) an H1 promoter operably linked to at least one nucleotide sequence encoding a CRISPR-Cas system guide RNA (gRNA), wherein the gRNA hybridizes with a target sequence of the DNA molecule; and b) a regulatory element operable in the eukaryotic cell operably linked to a nucleotide sequence encoding a Type-II Cas9 protein, wherein components (a) and (b) are located on the same or different vectors of the system, whereby the gRNA targets and hybridizes with the target sequence and the Cas9 protein cleaves the DNA molecule, and whereby expression of the one or more gene products is altered. In one aspect, the target sequence can be a target sequence that starts with any nucleotide, for example, N_{20}NGG. In some embodiments, the target sequence comprises the nucleotide sequence AN_{19}NGG. In some embodiments, the target sequence comprises the nucleotide sequence GN_{19}NGG. In some embodiments, the target sequence comprises the nucleotide sequence CN_{19}NGG. In some embodiments, the target sequence comprises the nucleotide sequence TN_{19}NGG. In some embodiments, the target sequence comprises the nucleotide sequence AN_{19}NGG or GN_{19}NGG. In another aspect, the Cas9 protein is codon optimized for expression in the cell. In yet another aspect, the Cas9 protein is codon optimized for expression in the eukaryotic cell. In a further aspect, the eukaryotic cell is a mammalian or human cell. In another aspect, the expression of the one or more gene products is decreased.

The presently disclosed subject matter also provides a method of altering expression of one or more gene products in a eukaryotic cell, wherein the cell comprises a DNA molecule encoding the one or more gene products, the method comprising introducing into the cell a non-naturally occurring CRISPR-Cas system comprising a vector comprising a bidirectional H1 promoter, wherein the bidirectional H1 promoter comprises: a) control elements that provide for transcription in one direction of at least one nucleotide sequence encoding a CRISPR-Cas system guide RNA (gRNA), wherein the gRNA hybridizes with a target sequence of the DNA molecule; and b) control elements that provide for transcription

in the opposite direction of a nucleotide sequence encoding a Type-II Cas9 protein, whereby the gRNA targets and hybridizes with the target sequence and the Cas9 protein cleaves the DNA molecule, and whereby expression of the one or more gene products is altered. In one aspect, the target sequence can be a target sequence that starts with any nucleotide, for example, $N_{20}NGG$. In some embodiments, the target sequence comprises the nucleotide sequence $AN_{19}NGG$. In some embodiments, the target sequence comprises the nucleotide sequence $GN_{19}NGG$. In some embodiments, the target sequence comprises the nucleotide sequence $CN_{19}NGG$. In some embodiments, the target sequence comprises the nucleotide sequence $TN_{19}NGG$. In another aspect, the target sequence comprises the nucleotide sequence $AN_{19}NGG$ or $GN_{19}NGG$. In another aspect, the Cas9 protein is codon optimized for expression in the cell. In yet another aspect, the Cas9 protein is codon optimized for expression in the eukaryotic cell. In a further aspect, the eukaryotic cell is a mammalian or human cell. In another aspect, the expression of the one or more gene products is decreased.

In some aspects, the presently disclosed subject matter provides methods comprising delivering one or more polynucleotides, such as or one or more vectors as described herein, one or more transcripts thereof, and/or one or proteins transcribed therefrom, to a host cell. In some aspects, the presently disclosed subject matter further provides cells produced by such methods, and organisms (such as animals, plants, or fungi) comprising or produced from such cells. In some embodiments, a CRISPR enzyme in combination with (and optionally complexed with) a guide sequence is delivered to a cell. Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids in mammalian cells or target tissues. Such methods can be used to administer nucleic acids encoding components of a CRISPR system to cells in culture, or in a host organism. Non-viral vector delivery systems include DNA plasmids, RNA (e.g. a transcript of a vector described herein), naked nucleic acid, and nucleic acid complexed with a delivery vehicle, such as a liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For a review of gene therapy procedures, see Anderson (1992) *Science* 256:808-813; Nabel and Felgner (1993) TIBTECH 11:211-217; Mitani and Caskey (1993) TIBTECH 11:162-166; Dillon (1993) TIBTECH 11:167-175; Miller (1992) *Nature* 357:455-460; Van Brunt (1998) *Biotechnology* 6(10): 1149-1154; Vigne (1995) *Restorative Neurology and Neuroscience* 8:35-36; Kremer and Perricaudet (1995) *British Medical Bulletin* 51(1):31-44; Haddada et

al. (1995) Current Topics in Microbiology and Immunology. Doerfler and Bohm (eds); and Yu et al. (1994) *Gene Therapy* 1:13-26.

Methods of non-viral delivery of nucleic acids include lipofection, nucleofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA. Lipofection is described in e.g., U.S. Pat. Nos. 5,049,386, 4,946,787; and 4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, WO 91/17424; WO 91/16024. Delivery can be to cells (e.g. in vitro or ex vivo administration) or target tissues (e.g. in vivo administration).

The preparation of lipid:nucleic acid complexes, including targeted liposomes such as immunolipid complexes, is well known to one of skill in the art (e.g., Crystal (1995) *Science* 270:404-410; Blaese et al. (1995) *Cancer Gene Ther.* 2:291-297: Behr et al. (1994) *Bioconjugate Chem.* 5:382-389; Remy et al. (1994) *Bioconjugate Chem.* 5:647-654; Gao et al. (1995) *Gene Therapy* 2:710-722; Ahmad et al. (1992) *Cancer Res.* 52:4817-4820; U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, and 4,946,787).

The use of RNA or DNA viral based systems for the delivery of nucleic acids take advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to patients (in vivo) or they can be used to treat cells in vitro, and the modified cells may optionally be administered to patients (ex vivo). Conventional viral based systems could include retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system would therefore depend on the target tissue. Retroviral vectors are comprised of cis-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum cis-

acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immuno deficiency virus (SIV), human immuno deficiency virus (HIV), and combinations thereof (e.g., Buchscher et al. (1992) *J. Virol.* 66:2731-2739; Johann et al. (1992) *J. Virol.* 66:1635-1640; Sommnerfelt et al. (1990) *J. Virol.* 176:58-59; Wilson et al. (1989) *J. Virol.* 63:2374-2378; Miller et al. (1991) *J. Virol.* 65:2220-2224; PCT/US94/05700). In applications where transient expression is preferred, adenoviral based systems may be used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors may also be used to transduce cells with target nucleic acids, e.g., in the *in vitro* production of nucleic acids and peptides, and for *in vivo* and *ex vivo* gene therapy procedures (e.g., West et al. (1987) *Virology* 160:38-47; U.S. Pat. No. 4,797,368; WO 93/24641; Kotin (1994) *Human Gene Therapy* 5:793-801; Muzyczka (1994) *J. Clin. Invest.* 94:1351. Construction of recombinant AAV vectors are described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin et al. (1985) *Mol. Cell. Biol.* 5:3251-3260; Tratschin et al. (1984) *Mol. Cell. Biol.* 4:2072-2081; Hermonat and Muzyczka (1984) *Proc. Natl. Acad. Sci. U.S.A.* 81:6466-6470; and Samulski et al. (1989) *J. Virol.* 63:03822-3828.

Packaging cells are typically used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and ψ2 cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by producing a cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the polynucleotide(s) to be expressed. The missing viral functions are typically supplied in trans by the packaging cell line. For example, AAV vectors used in gene therapy typically only possess ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely rep and cap, but lacking ITR sequences. The cell line may also be infected with adenovirus as a helper. The

helper virus promotes replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, e.g., heat treatment to which adenovirus is more sensitive than AAV. Additional methods for the delivery of nucleic acids to cells are known to those skilled in the art. See, for example, US20030087817, incorporated herein by reference.

In some embodiments, a host cell is transiently or non-transiently transfected with one or more vectors described herein. In some embodiments, a cell is transfected as it naturally occurs in a subject. In some embodiments, a cell that is transfected is taken from a subject. In some embodiments, the cell is derived from cells taken from a subject, such as a cell line. A wide variety of cell lines for tissue culture are known in the art. Examples of cell lines include, but are not limited to, C8161, CCRF-CEM, MOLT, mIMCD-3, NHDF, HeLa-S3, Huh1, Huh4, Huh7, HUVEC, HASMC, HEKn, HEKa, MiaPaCell, Panel, PC-3, TF1, CTLL-2, C1R, Rat6, CV1, RPTE, A10, T24, J82, A375, ARH-77, Calu1, SW480, SW620, SKOV3, SK-UT, CaCo2, P388D1, SEM-K2, WEHI-231, HB56, TIB55, Jurkat, J45.01, LRMB, Bcl-1, BC-3, IC21, DLD2, Raw264.7, NRK, NRK-52E, MRC5, MEF, Hep G2, HeLa B, HeLa T4, COS, COS-1, COS-6, COS-M6A, BS-C-1 monkey kidney epithelial, BALB/3T3 mouse embryo fibroblast, 3T3 Swiss, 3T3-L1, 132-d5 human fetal fibroblasts; 10.1 mouse fibroblasts, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A172, A20, A253, A431, A-549, ALC, B16, B35, BCP-1 cells, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO-IR, CHO-K1, CHO-K2, CHO-T, CHO Dhfr −/−, COR-L23, COR-L23/CPR, COR-L23/5010, COR-L23/R23, COS-7, COV-434, CML T1, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepa1c1c7, HL-60, HMEC, HT-29, Jurkat, JY cells, K562 cells, Ku812, KCL22, KG1, KYO1, LNCap, Ma-Mel 1-48, MC-38, MCF-7, MCF-10A, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCK II, MDCK II, MOR/0.2R, MONO-MAC 6, MTD-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, OPCN/OPCT cell lines, Peer, PNT-1A/PNT 2, RenCa, RIN-5F, RMA/RMAS, Saos-2 cells, Sf-9, SkBr3, T2, T-47D, T84, THP1 cell line, U373, U87, U937, VCaP, Vero cells, WM39, WT-49, X63, YAC-1, YAR, and transgenic varieties thereof. Cell lines are available from a variety of sources known to those with skill in the art (see, e.g., the American Type Culture Collection (ATCC) (Manassus, Va.)). In some embodiments, a cell

transfected with one or more vectors described herein is used to establish a new cell line comprising one or more vector-derived sequences. In some embodiments, a cell transiently transfected with the components of a CRISPR system as described herein (such as by transient transfection of one or more vectors, or transfection with RNA), and modified through the activity of a CRISPR complex, is used to establish a new cell line comprising cells containing the modification but lacking any other exogenous sequence. In some embodiments, cells transiently or non-transiently transfected with one or more vectors described herein, or cell lines derived from such cells are used in assessing one or more test compounds.

In some embodiments, one or more vectors described herein are used to produce a non-human transgenic animal. In some embodiments, the transgenic animal is a mammal, such as a mouse, rat, or rabbit. In certain embodiments, the organism or subject is a plant. Methods for producing transgenic animals are known in the art, and generally begin with a method of cell transfection, such as described herein.

In one aspect, the presently disclosed subject matter provides for methods of modifying a target polynucleotide in a eukaryotic cell, which may be in vivo, ex vivo or in vitro. In some embodiments, the method comprises sampling a cell or population of cells from a human or non-human animal, and modifying the cell or cells. Culturing may occur at any stage ex vivo. The cell or cells may even be re-introduced into the non-human animal.

In one aspect, the presently disclosed subject matter provides for methods of modifying a target polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a CRISPR complex to bind to the target polynucleotide to effect cleavage of the target polynucleotide thereby modifying the target polynucleotide, wherein the CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within the target polynucleotide.

In one aspect, the presently disclosed subject matter provides a method of modifying expression of a polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a CRISPR complex to bind to the polynucleotide such that the binding results in increased or decreased expression of the polynucleotide; wherein the CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within the polynucleotide.

In one aspect, the presently disclosed subject matter provides methods for using one or more elements of a CRISPR system. The CRISPR complex of the presently disclosed subject matter provides an effective means for modifying a target polynucleotide. The CRISPR complex of the presently disclosed subject matter has a wide variety of utility including modifying (e.g., deleting, inserting, translocating, inactivating, activating) a target polynucleotide in a multiplicity of cell types. As such the CRISPR complex of the presently disclosed subject matter has a broad spectrum of applications in, e.g., gene therapy, drug screening, disease diagnosis, and prognosis. An exemplary CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within the target polynucleotide.

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

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

responsible for the etiology of a disease. The transcribed or translated products may be known or unknown, and may be at a normal or abnormal level.

Embodiments of the presently disclosed subject matter also relate to methods and compositions related to knocking out genes, amplifying genes and repairing particular mutations associated with DNA repeat instability and neurological disorders (Robert D. Wells, Tetsuo Ashizawa, Genetic Instabilities and Neurological Diseases, Second Edition, Academic Press, Oct. 13, 2011-Medical). Specific aspects of tandem repeat sequences have been found to be responsible for more than twenty human diseases (McIvor et al. (2010) *RNA Biol.* 7(5):551-8). The CRISPR-Cas system may be harnessed to correct these defects of genomic instability.

In yet another aspect of the presently disclosed subject matter, the CRISPR-Cas system may be used to correct ocular defects that arise from several genetic mutations further described in Traboulsi, ed. (2012) Genetic Diseases of the Eye, Second Edition, Oxford University Press.

Several further aspects of the presently disclosed subject matter relate to correcting defects associated with a wide range of genetic diseases. For example, genetic brain diseases may include but are not limited to Adrenoleukodystrophy, Agenesis of the Corpus Callosum, Aicardi Syndrome, Alpers' Disease. Alzheimer's Disease, Barth Syndrome, Batten Disease, CADASIL, Cerebellar Degeneration, Fabry's Disease, Gerstmann-Straussler-Scheinker Disease, Huntington's Disease and other Triplet Repeat Disorders, Leigh's Disease, Lesch-Nyhan Syndrome, Menkes Disease, Mitochondrial Myopathies and NINDS Colpocephaly.

In some embodiments, the condition may be neoplasia. In some embodiments, the condition may be Age-related Macular Degeneration. In some embodiments, the condition may be a Schizophrenic Disorder. In some embodiments, the condition may be a Trinucleotide Repeat Disorder. In some embodiments, the condition may be Fragile X Syndrome. In some embodiments, the condition may be a Secretase Related Disorder. In some embodiments, the condition may be a Prion—related disorder. In some embodiments, the condition may be ALS. In some embodiments, the condition may be a drug addiction. In some embodiments, the condition may be Autism. In some embodiments, the condition may be Alzheimer's Disease. In some embodiments, the condition may be inflammation. In some

embodiments, the condition may be Parkinson's Disease.

Examples of proteins associated with Parkinson's disease include but are not limited to α-synuclein, DJ-1, LRRK2, PINK1, Parkin, UCHL1, Synphilin-1, and NURR1.

Examples of addiction-related proteins may include ABAT for example.

Examples of inflammation-related proteins may include the monocyte chemoattractant protein-1 (MCP1) encoded by the Ccr2 gene, the C—C chemokine receptor type 5 (CCR5) encoded by the Ccr5 gene, the IgG receptor IIB (FCGR2b, also termed CD32) encoded by the Fcgr2b gene, or the Fc epsilon R1g (FCER1g) protein encoded by the Fcer1g gene, for example.

Examples of cardiovascular disease associated proteins may include IL1B (interleukin 1, beta), XDH (xanthine dehydrogenase), TP53 (tumor protein p53), PTGIS (prostaglandin 12 (prostacyclin) synthase), MB (myoglobin), IL4 (interleukin 4), ANGPT1 (angiopoietin 1), ABCG8 (ATP-binding cassette, sub-family G (WHITE), member 8), or CTSK (cathepsin K), for example.

Examples of Alzheimer's disease associated proteins may include the very low density lipoprotein receptor protein (VLDLR) encoded by the VLDLR gene, the ubiquitin-like modifier activating enzyme 1 (UBA1) encoded by the UBA1 gene, or the NEDD8-activating enzyme E1 catalytic subunit protein (UBE1C) encoded by the UBA3 gene, for example.

Examples of proteins associated Autism Spectrum Disorder may include the benzodiazapine receptor (peripheral) associated protein 1 (BZRAP1) encoded by the BZRAP1 gene, the AF4/FMR2 family member 2 protein (AFF2) encoded by the AFF2 gene (also termed MFR2), the fragile X mental retardation autosomal homolog 1 protein (FXR1) encoded by the FXR1 gene, or the fragile X mental retardation autosomal homolog 2 protein (FXR2) encoded by the FXR2 gene, for example.

Examples of proteins associated Macular Degeneration may include the ATP-binding cassette, sub-family A (ABC1) member 4 protein (ABCA4) encoded by the ABCR gene, the apolipoprotein E protein (APOE) encoded by the APOE gene, or the chemokine (C—C motif) Ligand 2 protein (CCL2) encoded by the CCL2 gene, for example.

Examples of proteins associated Schizophrenia may include NRG1, ErbB4, CPLX1, TPH1, TPH2, NRXN1, GSK3A, BDNF, DISC1, GSK3B, and combinations thereof.

Examples of proteins involved in tumor suppression may include ATM (ataxia telangiectasia mutated), ATR (ataxia telangiectasia and Rad3 related), EGFR (epidermal

growth factor receptor), ERBB2 (v-erb-b2 erythroblastic leukemia viral oncogene homolog 2), ERBB3 (v-erb-b2 erythroblastic leukemia viral oncogene homolog 3), ERBB4 (v-erb-b2 erythroblastic leukemia viral oncogene homolog 4), Notch 1, Notch2, Notch 3, or Notch 4, for example.

Examples of proteins associated with a secretase disorder may include PSENEN (presenilin enhancer 2 homolog (*C. elegans*)), CTSB (cathepsin B), PSEN1 (presenilin 1), APP (amyloid beta (A4) precursor protein), APH1B (anterior pharynx defective 1 homolog B (*C. elegans*)), PSEN2 (presenilin 2 (Alzheimer disease 4)), or BACE1 (beta-site APP-cleaving enzyme 1), for example.

Examples of proteins associated with Amyotrophic Lateral Sclerosis may include SOD1 (superoxide dismutase 1), ALS2 (amyotrophic lateral sclerosis 2), FUS (fused in sarcoma), TARDBP (TAR DNA binding protein), VAGFA (vascular endothelial growth factor A), VAGFB (vascular endothelial growth factor B), and VAGFC (vascular endothelial growth factor C), and any combination thereof.

Examples of proteins associated with prion diseases may include SOD1 (superoxide dismutase 1), ALS2 (amyotrophic lateral sclerosis 2), FUS (fused in sarcoma), TARDBP (TAR DNA binding protein), VAGFA (vascular endothelial growth factor A), VAGFB (vascular endothelial growth factor B), and VAGFC (vascular endothelial growth factor C), and any combination thereof.

Examples of proteins related to neurodegenerative conditions in prion disorders may include A2M (Alpha-2-Macroglobulin), AATF (Apoptosis antagonizing transcription factor), ACPP (Acid phosphatase prostate), ACTA2 (Actin alpha 2 smooth muscle aorta), ADAM22 (ADAM metallopeptidase domain), ADORA3 (Adenosine A3 receptor), or ADRA1D (Alpha-1D adrenergic receptor for Alpha-1D adrenoreceptor), for example.

Examples of proteins associated with immunodeficiency may include A2M [alpha-2-macroglobulin]; AANAT [arylalkylamine N-acetyltransferase]; ABCA1 [ATP-binding cassette, sub-family A (ABC1), member 1]; ABCA2 [ATP-binding cassette, sub-family A (ABC1), member 2]; or ABCA3 [ATP-binding cassette, sub-family A (ABC1), member 3]; for example.

Examples of proteins associated with Trinucleotide Repeat Disorders include AR (androgen receptor), FMR1 (fragile X mental retardation 1), HTT (huntingtin), or DMPK (dystrophia myotonica-protein kinase), FXN (frataxin), ATXN2 (ataxin 2). for example.

Examples of proteins associated with Neurotransmission Disorders include SST (somatostatin), NOS1 (nitric oxide synthase 1 (neuronal)), ADRA2A (adrenergic, alpha-2A-, receptor), ADRA2C (adrenergic, alpha-2C-, receptor), TACR1 (tachykinin receptor 1), or HTR2c (5-hydroxytryptamine (serotonin) receptor 2C), for example.

Examples of neurodevelopmental-associated sequences include A2BP1 (ataxin 2-binding protein 1), AADAT (aminoadipate aminotransferase), AANAT (arylalkylamine N-acetyltransferase), ABAT (4-aminobutyrate aminotransferase), ABCA1 (ATP-binding cassette, sub-family A (ABC1), member 1), or ABCA13 (ATP-binding cassette, sub-family A (ABC1), member 13), for example.

Further examples of preferred conditions treatable with the present system include may be selected from: Aicardi-Goutières Syndrome; Alexander Disease; Allan-Herndon-Dudley Syndrome; POLG-Related Disorders; Alpha-Mannosidosis (Type II and III); Alström Syndrome; Angelman; Syndrome; Ataxia-Telangiectasia; Neuronal Ceroid-Lipofuscinoses; Beta-Thalassemia; Bilateral Optic Atrophy and (Infantile) Optic Atrophy Type 1; Retinoblastoma (bilateral); Canavan Disease; Cerebrooculofacioskeletal Syndrome 1 (COFS1); Cerebrotendinous Xanthomatosis; Cornelia de Lange Syndrome; MAPT-Related Disorders; Genetic Prion Diseases; Dravet Syndrome; Early-Onset Familial Alzheimer Disease; Friedreich Ataxia [FRDA]; Fryns Syndrome; Fucosidosis; Fukuyama Congenital Muscular Dystrophy; Galactosialidosis; Gaucher Disease; Organic Acidemias; Hemophagocytic Lymphohistiocytosis; Hutchinson-Gilford Progeria Syndrome; Mucolipidosis II; Infantile Free Sialic Acid Storage Disease; PLA2G6-Associated Neurodegeneration; Jervell and Lange-Nielsen Syndrome; Junctional Epidermolysis Bullosa; Huntington Disease; Krabbe Disease (Infantile); Mitochondrial DNA-Associated Leigh Syndrome and NARP; Lesch-Nyhan Syndrome; LIS1-Associated Lissencephaly; Lowe Syndrome; Maple Syrup Urine Disease; MECP2 Duplication Syndrome; ATP7A-Related Copper Transport Disorders; LAMA2-Related Muscular Dystrophy; Arylsulfatase A Deficiency; Mucopolysaccharidosis Types I, II or III; Peroxisome Biogenesis Disorders, Zellweger Syndrome Spectrum; Neurodegeneration with Brain Iron Accumulation Disorders; Acid Sphingomyelinase Deficiency; Niemann-Pick Disease Type C; Glycine Encephalopathy; ARX-Related Disorders; Urea Cycle Disorders; COL1A 1/2-Related Osteogenesis Imperfecta; Mitochondrial DNA Deletion Syndromes; PLP1-Related Disorders; Perry Syndrome; Phelan-McDermid Syndrome; Glycogen Storage Disease Type II (Pompe Disease) (Infantile); MAPT-Related Disorders; MECP2-Related Disorders;

Rhizomelic Chondrodysplasia Punctata Type 1; Roberts Syndrome; Sandhoff Disease; Schindler Disease—Type 1; Adenosine Deaminase Deficiency; Smith-Lemli-Opitz Syndrome; Spinal Muscular Atrophy, Infantile-Onset Spinocerebellar Ataxia; Hexosaminidase A Deficiency; Thanatophoric Dysplasia Type 1; Collagen Type VI-Related Disorders; Usher Syndrome Type I; Congenital Muscular Dystrophy; Wolf-Hirschhorn Syndrome; Lysosomal Acid Lipase Deficiency; and Xeroderma Pigmentosum.

II. METHODS FOR TREATING NEURODEGENERATIVE DISEASES

The presently disclosed subject matter also provides methods for treating a disease selected from the group consisting of retinal dystrophy, corneal dystrophy, microsatellite expansion diseases, cancer, and neurodegenerative diseases in a subject in need thereof, the method comprising:(a) providing a non-naturally occurring nuclease system (e.g., CRISPR associated (Cas) 9 (CRISPR-Cas9, non-Cas9 CRISPR systems, CRISPR-Cpf-1 system, and the like) comprising one or more vectors comprising: a) a promoter (e.g., orthogolous H1 promoter or 7sk) operably linked to at least one nucleotide sequence encoding a nuclease system guide RNA (gRNA), wherein the gRNA hybridizes with a target sequence of a DNA molecule in a cell, and wherein the DNA molecule encodes one or more gene products expressed in the cell; and b) a regulatory element operable in a cell operably linked to a nucleotide sequence encoding a genome-targeted nuclease (e.g., modified Cas9 protein), wherein components (a) and (b) are located on the same or different vectors of the system, wherein the gRNA targets and hybridizes with the target sequence and the nuclease cleaves the DNA molecule to alter expression of the one or more gene products; and (b) administering to the subject an effective amount of the system. In some embodiments, the system is packaged into a single adeno-associated virus (AAV) particle. In some embodiments, the system inactivates one or more gene products. In some embodiments, the system excises at least one gene mutation. In some embodiments, the H1 promoter comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 12 or 32. In some embodiments, the orthologous H1 promoter is derived from mouse or rat. In some embodiments, the orthologous H1 promoter comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NOs: 84-119. In some embodiments, the orthologous H1 promoter comprises a nucleotide sequences set forth in the group consisting of SEQ ID NOs: 84-119. In some

embodiments, the H1 promoter comprises: a) control elements that provide for transcription in one direction of at least one nucleotide sequence encoding a gRNA; and b) control elements that provide for transcription in the opposite direction of a nucleotide sequence encoding a genome-targeted nuclease. In some embodiments, the promoter is the human, mouse, or rat 7sk promoter. In some embodiments, the 7sk is selected from the group consisting of 7sk1 (variant 1 of 7sk which uses the endogenous GSTA4 5'UTR), 7sk2 (variant 2 of 7sk which uses just a kozak sequence in the 5'UTR), and 7sk3 (variant of 7sk which uses the beta globin 5'UTR). In some embodiments, the human 7sk1 comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 3. In some embodiments, the mouse 7sk1 comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to SEQ ID NO: 6. In some embodiments, the human 7sk2 comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to SEQ ID NO: 4. In some embodiments, the mouse 7sk2 comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to SEQ ID NO: 7. In some embodiments, the human 7sk3 comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to SEQ ID NO: 5. In some embodiments, the mouse 7sk3 comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to SEQ ID NO: 8. In some embodiments, the orthologous H1 promoter or 7sk promoter has at least one modification. In some embodiments, the at least one modification of the promoter comprises an element that allows conditional regulation. In some embodiments, the element is a tet-responsive promoter. In some embodiments, the tet-response promoter comprises a Tet repressor (TetR) and Tet operator (TetO) engineered into the H1 promoter. In some embodiments, the TetR comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to SEQ ID NO: 14. In some embodiments, the TetO comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 11. In some embodiments, the H1-TetO comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 13. In some embodiments, the at least one modification of the promoter comprises a site that allows auto-regulation. In some embodiments, the auto-regulation site comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 33 or 34. In some

embodiments, the nuclease system further comprising at least one terminator sequence. In some embodiments, the at least one terminator sequence is selected from the group consisting of SV40 or synthetic poly A (SPA) sequences. In some embodiments, the terminator sequences is a SV40 120 (SEQ ID NO: 123) or 240 (SEQ ID NO: 122) base pair sequence. In some embodiments, the terminator sequences is a SPA 49 base pair equence. In some embodiments, the nuclease system further comprises 5' untranslated region (5'UTR) sequences. In some embodiments, the nuclease system further comprises a Kozak sequence. In some embodiments, the Kozak sequence comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to SEQ ID NO: 1 or 2. In some embodiments, the nuclease system further comprises a RNA sequence that mediates cap-independent initiation of translation. In some embodiments, the RNA sequence is selected from the group consisting of 6.947 or 6.967. In some embodiments, the Cas9 comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 61. In some embodiments, the Cas9 comprises at least one modification. In some embodiments, the at least one modification in the Cas9 comprises an alteration in the cleaving sequence. In some embodiments, the nuclease is seleted from the group consisting of Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 and Csx12), Cas10, Cas13a, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, Cpf1, C2c1, C2c2, and C2c3. In some embodiments, the alteration in the cleaving sequence selected from the group consisting of T2A, P2A, E2A, and F2A. In some embodiments, the T2A comprises an amino acid sequence. having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 36. In some embodiments, the P2A comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 35. In some embodiments, the P2A comprises an amino acid sequence having the nucleotide sequence set forth in SEQ ID NO: 35. In some embodiments, the E2A comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 37. In some embodiments, the F2A comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 38. In some

embodiments, the at least one modification in the Cas9 comprises a codon optimized for expression in the cell. In some embodiments, the Cas9 further comprises a linker sequence operably fused in frame to a cell-cycle dependent protein (Cas9 fusion). In some embodiments, the cell cycle-dependent protein is selected from the group consisting of APE2, ATR, BRCA1, Chk1, Cdc5, Cdc6, Cdc7, Cdc45, Cdt1, CSA, CSB, Ctf18, DDB1, DDB2, DNA2, DUT, Elg1, EndoV, Esp1, Exonuclease1, FBH1, FEN1, Geminin, Hus1, KNTC2 (NDC80), Ku80, Ligase1, Mad2, MBD4, Mcm3, Mcm4, Mcm5, Mcm6, Mcm7, Mcm8, Mcm10, MGMT, MLH3, Mms4, MPG, MSH2, Mus81, NBS1, NEIL2, NEIL3, NTH1, Orc1, Orc3, PARP1, PCNA, Pif1, PMS1, PMS2, PNK, Pola p180, Pola p70, Pola Spp1 (Prim2a), Polb, Pold p125, Pole Dpb3, Pole Dpb4, Pole Pol2, Poli, Poll, Polm, Psf1, Psf2, Psf3, Rad1, Rad18, Rad23A, Rad23B, Rad51, Rad51D, Rad54, Rad6A, RPA34, RPA70, Scc1, Scc3, Sir2, SIRT1 (Sirtuin), TDG, TDP1, TIMELESS, Tin2, Topoisomerase I, Topoisomerase IIIa, Topoisomerase IIIb, Ubc13, UNG, XAB2, XPC, XPF, XPG, Xrcc2, and XRCC4. In some embodiments, the cell cycle-dependent protein is Geminin. In some embodiments, the cell cycle-dependent protein is human Geminin. In some embodiments, the human Geminin comprises the amino acids from positions 1-110 (hGem(1-110)). In some embodiments, the hGem(1-110) comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the amimo acid sequence set forth in SEQ ID NO: 19. In some embodiments, the cell cycle-dependent protein is Cdt1. In some embodiments, the cell cycle-dependent protein is human Cdt1. In some embodiments, the human Cdt1 comprises amino acids from positions 30-120 (hCdt1(30-120)). In some embodiments, the hCdt1(30-120) comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the amino acid sequence set forth in SEQ ID NO: 18. In some embodiments, the Cas9 fusion comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the amino acid sequence selected from the group consisting of SEQ ID NO: 20-27. In some embodiments, the Cas9 is operably fused in frame to a ubiquitin protein (Ub-Cas9). In some embodiments, the Ub-Cas9 at least one N-terminal modification. In some embodiments, the N-terminal modified Ub-Cas9 comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the amino acid sequence selected from the group consisting of SEQ ID NO: 39-58. In some embodiments, the ubiquitin protein comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 60. In some embodiments, the

nuclease system further comprises a SaCas9 nickase. In some embodiments, the nuclease system further comprises a donor template sequence. In some embodiments, the at least one vector comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 62 or 67. In some embodiments, the donor template sequence corrects at least one gene mutation. In some embodiments, the at least one gene mutation is *rd10* or *rd12*. In some embodiments, the at least one vector comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 63 or 64. In some embodiments, the at least one vector comprises a nucleotide sequence having at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identity to the nucleotide sequence set forth in SEQ ID NO: 65 or 66. In some embodiments, the promoter is operably linked to at least one, two, three, four, five, six, seven, eight, nine, or ten gRNA. In some embodiments, the target sequence comprises the nucleotide sequence $AN_{19}NGG$, $GN_{19}NGG$, $CN_{19}NGG$, or $TN_{19}NGG$. In some embodiments, the cell is a eukaryotic or or non-eukaryotic cell. In some embodiments, the eukaryotic cell is a mammalian or human cell. In some embodiments, the eukaryotic cell is a retinal photoreceptor cell. In some embodiments, the one or more gene products are rhodopsin. In some embodiments, the expression of the one or more gene products is decreased. In some embodiments, the system inactivates one or more gene products. In some embodiments, the nuclease system excises at least one gene mutation. In some embodiments, the expression of the one or more gene products is decreased. In some embodiments, the cell is a retinal ganglion cell. In some embodiments, the eukaryotic cell is a cancerous cell. In some embodiments, cell proliferation is inhibited or reduced in the cancerous cell. In some embodiments, the apoptosis is enhanced or increased in the cancerous cell. In some embodiments, the disease is corneal dystrophy. In some embodiments, the disease is a retinal dystrophy. In some embodiments, the retinal dystrophy is selected from the group consisting of Leber's congenital amaurosis (LCA), retinitis pigmentosa (RP), and glaucoma. In some embodiments, the corneal dystrophy is selected from the group consisting of Epithelial Basement Membrane Dystrophy, Epithelial Recurrent Erosion Dystrophies, Subepithelial Mucinous Corneal Dystrophy, Meesmann Corneal Dystrophy, Lisch Epithelial Corneal Dystrophy, Gelatinous Drop-like Corneal Dystrophy, Reis–Bucklers Corneal Dystrophy, Thiel–Behnke Corneal Dystrophy, Lattice Corneal Dystrophy, Type 1 (Classic), Lattice Corneal Dystrophy, Type 2, Lattice Corneal Dystrophy, Type III, Lattice Corneal

Dystrophy, Type IIIA, Lattice Corneal Dystrophy, Type I/IIIA, Lattice Corneal Dystrophy, Type IV, Polymorphic (Corneal) Amyloidosis, Granular Corneal Dystrophy, Type 1, Granular Corneal Dystrophy, Type 2, Macular Corneal Dystrophy, Schnyder Corneal Dystrophy, Congenital Stromal Corneal Dystrophy, Fleck Corneal Dystrophy, Posterior Amorphous Corneal Dystrophy, Central Cloudy Dystrophy of Francois, Pre-Descemet Corneal Dystrophy, Fuchs Endothelial Corneal Dystrophy, Posterior Polymorphous Corneal Dystrophy, Congenital Hereditary Endothelial Dystrophy, and X-linked Endothelial Corneal Dystrophy. In some embodiments, the microsatellite expansion diseases is selected from the group consisting of Blepharophimosis, ptosis and epicanthus inversus syndactyly, Cleidocranial dysplasia, Congenital central hypoventilation syndrome, Haddad syndrome DM (Myotonic dystrophy), FRAXA (Fragile X syndrome), FRAXE (Fragile XE mental retardation), FRDA (Friedreich's ataxia), Fuchs' Endothelial Corneal Dystrophy, FXTAS (Fragile X-associated tremor/ataxia syndrome), Hand-foot-genital syndrome, HD (Huntington's disease), Holoprosencephaly, Mental retardation with growth hormone deficiency, Mental retardation, epilepsy, West syndrome, Partington syndrome, Oculopharyngeal muscular dystrophy, SBMA (Spinal and bulbar muscular atrophy), SCA1 (Spinocerebellar ataxia Type 1), SCA12 (Spinocerebellar ataxia Type 12), SCA17 (Spinocerebellar ataxia Type 17), SCA2 (Spinocerebellar ataxia Type 2), SCA3 (Spinocerebellar ataxia Type 3 or Machado-Joseph disease), SCA6 (Spinocerebellar ataxia Type 6), SCA7 (Spinocerebellar ataxia Type 7), SCA8 (Spinocerebellar ataxia Type 8), and Synpolydactyly. In some embodiments, administering to the subject occurs by implantation, injection, or virally. In some embodiments, administering to the subject occurs by subretinal injection. In some embodiments, administering to the subject occurs topically, intravascularly, intradermally, transdermally, parenterally, intravenously, intramuscularly, intranasally, subcutaneously, regionally, percutaneously, intratracheally, intraperitoneally, intraarterially, intravesically, intratumorally, peritumorally, inhalationly, systematically, perfusionly, lavagely, directly via injection, or orally via administration and formulation. In some embodiments, administering to the subject occurs topically to the surface of the eye. In some embodiments, administering to the subject occurs on or outside the cornea, sclera, to the intraocular, subconjunctival, sub-tenon, or retrobulbar space, or in or around the eyelids. In some embodiments, the subject is treated with at least one additional anti-cancer agent. In some embodiments, the anti-cancer agent is selected from the group consisting of paclitaxel, cisplatin, topotecan, gemcitabine, bleomycin, etoposide, carboplatin, docetaxel,

doxorubicin, topotecan, cyclophosphamide, trabectedin, olaparib, tamoxifen, letrozole, and bevacizumab. In some embodiments, the subject is treated with at least one additional anti-cancer therapy. In some embodiments, the anti-cancer therapy is radiation therapy, chemotherapy, or surgery. In some embodiments, the cancer is a solid tumor. In some embodiments, the cancer is selected from the group consisting of brain cancer, gastrointestinal cancer, oral cancer, breast cancer, ovarian cancer, prostate cancer, pancreatic cancer, lung cancer, liver cancer, throat cancer, stomach cancer, and kidney cancer. In some embodiments, the cancer is brain cancer. In some embodiments, the systematic administration is selected from the group consisting of oral, intravenous, intradermal, intraperitoneal, subcutaneous, and intramuscular administration. In some embodiments, the subject is a mammal. In some embodiments, the mammal is human.

In some embodiments, the presently disclosed subject matter provides a method for treating an ocular neurodegenerative disease in a subject in need thereof, the method comprising: (a) providing a non-naturally occurring CRISPR-Cas system comprising one or more vectors comprising: i) an H1 promoter operably linked to at least one nucleotide sequence encoding a CRISPR-Cas system guide RNA (gRNA), wherein the gRNA hybridizes with a target sequence of a DNA molecule in a cell of the subject, and wherein the DNA molecule encodes one or more gene products expressed in the cell; and ii) a regulatory element operable in a cell operably linked to a nucleotide sequence encoding a Cas9 protein, wherein components (i) and (ii) are located on the same or different vectors of the system, wherein the gRNA targets and hybridizes with the target sequence and the Cas9 protein cleaves the DNA molecule to alter expression of the one or more gene products; and (b) administering to the subject an effective amount of the system.

By "neurodegenerative disease, disorder, or condition" is meant a disease, disorder, or condition (including a neuropathy) associated with degeneration or dysfunction of neurons or other neural cells, such as retinal photoreceptor cells. A neurodegenerative disease, disorder, or condition can be any disease, disorder, or condition in which decreased function or dysfunction of neurons, or loss or neurons or other neural cells, can occur.

Such diseases, disorders, or conditions include, but are not limited to, glaucoma, and neurodegenerative diseases, disorders, or conditions of the nervous systems, such as or associated with amyotrophic lateral sclerosis (ALS), trigeminal neuralgia, glossopharyngeal neuralgia, Bell's Palsy, myasthenia gravis, muscular dystrophy, progressive muscular atrophy, primary lateral sclerosis (PLS), pseudobulbar palsy, progressive bulbar palsy,

spinal muscular atrophy, inherited muscular atrophy, invertebrate disk syndromes, cervical spondylosis, plexus disorders, thoracic outlet destruction syndromes, peripheral neuropathies, prophyria, Alzheimer's disease, Huntington's disease, Parkinson's disease, Parkinson's-plus diseases, multiple system atrophy, progressive supranuclear palsy, corticobasal degeneration, dementia with Lewy bodies, frontotemporal dementia, demyelinating diseases, Guillain-Barre syndrome, multiple sclerosis, Charcot-Marie-Tooth disease, prion diseases, Creutzfeldt-Jakob disease, Gerstmann-Straussler-Scheinker syndrome (GSS), fatal familial insomnia (FFI), bovine spongiform encephalopathy (BSE), Pick's disease, epilepsy, and AIDS demential complex.

Other neurodegenerative diseases, disorders, or conditions of the nervous systems, such as or associated with alcoholism, Alexander's disease, Alper's disease, ataxia telangiectasia, Batten disease (also known as Spielmeyer-Vogt-Sjogren-Batten disease), Canavan disease, Cockayne syndrome, diabetic neuropathy, frontotemporal lobar degeneration, HIV-associated dementia, Kennedy's disease, Krabbe's disease, neuroborreliosis, Machado-Joseph disease (Spinocerebellar ataxia type 3), wet or dry macular degeneration, Niemann Pick disease, Pelizaeus-Merzbacher Disease, photoreceptor degenerative diseases, such as retinitis pigmentosa and associated diseases, Refsum's disease, Sandhoff's disease, Schilder's disease, subacute combined degeneration of spinal cord secondary to pernicious anemia, Spielmeyer-Vogt- Sjogren-Batten disease (also known as Batten disease), spinocerebellar ataxia (multiple types with varying characteristics), Steele-Richardson-Olszewski disease, and tabes dorsalis.

Examples of ocular-related neurodegeneration include, but are not limited to, glaucoma, lattice dystrophy, retinitis pigmentosa, age-related macular degeneration (AMD), photoreceptor degeneration associated with wet or dry AMD, other retinal degeneration such as retinitis pigmentosa (RP), optic nerve drusen, optic neuropathy, and optic neuritis, such as optic neuritis resulting from multiple sclerosis. In some embodiments, the ocular neurodegenerative disease is selected from the group consisting of glaucoma, retinal degeneration, and age-related macular degeneration. In some embodiments, the ocular neurodegenerative disease is retinitis pigmentosa (RP).

Non-limiting examples of different types of glaucoma that can be prevented or treated according to the presently disclosed subject matter include primary glaucoma (also known as primary open-angle glaucoma, chronic open-angle glaucoma, chronic simple glaucoma, and glaucoma simplex), low-tension glaucoma, primary angle- closure glaucoma

(also known as primary closed-angle glaucoma, narrow-angle glaucoma, pupil-block glaucoma, and acute congestive glaucoma), acute angle-closure glaucoma, chronic angle-closure glaucoma, intermittent angle-closure glaucoma, chronic open-angle closure glaucoma, pigmentary glaucoma, exfoliation glaucoma (also known as pseudoexfoliative glaucoma or glaucoma capsulare), developmental glaucoma (e.g., primary congenital glaucoma and infantile glaucoma), secondary glaucoma (e.g., inflammatory glaucoma (e.g., uveitis and Fuchs heterochromic iridocyclitis)), phacogenic glaucoma (e.g., angle-closure glaucoma with mature cataract, phacoanaphylactic glaucoma secondary to rupture of lens capsule, phacolytic glaucoma due to phacotoxic meshwork blockage, and subluxation of lens), glaucoma secondary to intraocular hemorrhage (e.g., hyphema and hemolytic glaucoma, also known as erythroclastic glaucoma), traumatic glaucoma (e.g., angle recession glaucoma, traumatic recession on anterior chamber angle, postsurgical glaucoma, aphakic pupillary block, and ciliary block glaucoma), neovascular glaucoma, drug-induced glaucoma (e.g., corticosteroid induced glaucoma and alpha- chymotrypsin glaucoma), toxic glaucoma, and glaucoma associated with intraocular tumors, retinal detachments, severe chemical burns of the eye, and iris atrophy. In certain embodiments, the neurodegenerative disease, disorder, or condition is a disease, disorder, or condition that is not associated with excessive angiogenesis, for example, a glaucoma that is not neovascular glaucoma.

As used herein, the term "disorder" in general refers to any condition that would benefit from treatment with a compound against one of the identified targets, or pathways, including any disease, disorder, or condition that can be treated by an effective amount of a compound against one of the identified targets, or pathways, or a pharmaceutically acceptable salt thereof.

As used herein, the term “treating” can include reversing, alleviating, inhibiting the progression of, preventing or reducing the likelihood of the disease, disorder, or condition to which such term applies, or one or more symptoms or manifestations of such disease, disorder or condition (e.g., a disease or disorder that causes dysfunction and/or death of retinal photoreceptor cells). In some embodiments, the treatment reduces the dysfunction and/or death of retinal photoreceptor cells. For example, the treatment can reduce the dysfunction and/or death of retinal photoreceptor cells by at least 5%, 10%, 15%, 20%, 25%, 30%, 33%, 35%, 40%, 45%, 50%, 55%, 60%, 66%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more as compared to the dysfunction and/or death of retinal photoreceptor cells in a subject before undergoing treatment or in a

subject who does not undergo treatment. In some embodiments, the treatment completely inhibits dysfunction and/or death of retinal photoreceptor cells in the subject. As used herein, a "retinal photoreceptor cell" is a specialized type of neuron found in the retina that is capable of phototransduction. In some embodiments, at least one gene product is rhodopsin.

In some embodiments, the system is packaged into a single adeno-associated virus (AAV) particle before administering to the subject. In some embodiments, administering to the subject occurs by subretinal injection. The treatment, administration, or therapy can be consecutive or intermittent. Consecutive treatment, administration, or therapy refers to treatment on at least a daily basis without interruption in treatment by one or more days. Intermittent treatment or administration, or treatment or administration in an intermittent fashion, refers to treatment that is not consecutive, but rather cyclic in nature. Treatment according to the presently disclosed methods can result in complete relief or cure from a disease, disorder, or condition, or partial amelioration of one or more symptoms of the disease, disease, or condition, and can be temporary or permanent. The term "treatment" also is intended to encompass prophylaxis, therapy and cure.

The term "effective amount" or "therapeutically effective amount" refers to the amount of an agent that is sufficient to effect beneficial or desired results. The therapeutically effective amount may vary depending upon one or more of: the subject and disease condition being treated, the weight and age of the subject, the severity of the disease condition, the manner of administration and the like, which can readily be determined by one of ordinary skill in the art. The term also applies to a dose that will provide an image for detection by any one of the imaging methods described herein. The specific dose may vary depending on one or more of: the particular agent chosen, the dosing regimen to be followed, whether it is administered in combination with other compounds, timing of administration, the tissue to be imaged, and the physical delivery system in which it is carried.

The terms "subject" and "patient" are used interchangeably herein. The subject treated by the presently disclosed methods in their many embodiments is desirably a human subject, although it is to be understood that the methods described herein are effective with respect to all vertebrate species, which are intended to be included in the term "subject." Accordingly, a "subject" can include a human subject for medical purposes, such as for the treatment of an existing condition or disease or the prophylactic treatment for preventing

the onset of a condition or disease, or an animal subject for medical, veterinary purposes, or developmental purposes. Suitable animal subjects include mammals including, but not limited to, primates, e.g., humans, monkeys, apes, and the like; bovines, e.g., cattle, oxen, and the like; ovines, e.g., sheep and the like; caprines, e.g., goats and the like; porcines, e.g., pigs, hogs, and the like; equines, e.g., horses, donkeys, zebras, and the like; felines, including wild and domestic cats; canines, including dogs; lagomorphs, including rabbits, hares, and the like; and rodents, including mice, rats, and the like. An animal may be a transgenic animal. In some embodiments, the subject is a human including, but not limited to, fetal, neonatal, infant, juvenile, and adult subjects. Further, a "subject" can include a patient afflicted with or suspected of being afflicted with a condition or disease.

IV. GENERAL DEFINITIONS

Although specific terms are employed herein, they are used in a generic and descriptive sense only and not for purposes of limitation. Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this presently described subject matter belongs.

Following long-standing patent law convention, the terms "a," "an," and "the" refer to "one or more" when used in this application, including the claims. Thus, for example, reference to "a subject" includes a plurality of subjects, unless the context clearly is to the contrary (e.g., a plurality of subjects), and so forth.

Throughout this specification and the claims, the terms "comprise," "comprises," and "comprising" are used in a non-exclusive sense, except where the context requires otherwise. Likewise, the term "include" and its grammatical variants are intended to be non-limiting, such that recitation of items in a list is not to the exclusion of other like items that can be substituted or added to the listed items.

For the purposes of this specification and appended claims, unless otherwise indicated, all numbers expressing amounts, sizes, dimensions, proportions, shapes, formulations, parameters, percentages, parameters, quantities, characteristics, and other numerical values used in the specification and claims, are to be understood as being modified in all instances by the term "about" even though the term "about" may not expressly appear with the value, amount or range. Accordingly, unless indicated to the contrary, the numerical parameters set forth in the following specification and attached claims are not and need not be exact, but may be approximate and/or larger or smaller as

desired, reflecting tolerances, conversion factors, rounding off, measurement error and the like, and other factors known to those of skill in the art depending on the desired properties sought to be obtained by the presently disclosed subject matter. For example, the term "about," when referring to a value can be meant to encompass variations of, in some embodiments, ± 100% in some embodiments ± 50%, in some embodiments ± 20%, in some embodiments ± 10%, in some embodiments ± 5%, in some embodiments ±1%, in some embodiments ± 0.5%, and in some embodiments ± 0.1% from the specified amount, as such variations are appropriate to perform the disclosed methods or employ the disclosed compositions.

Further, the term "about" when used in connection with one or more numbers or numerical ranges, should be understood to refer to all such numbers, including all numbers in a range and modifies that range by extending the boundaries above and below the numerical values set forth. The recitation of numerical ranges by endpoints includes all numbers, e.g., whole integers, including fractions thereof, subsumed within that range (for example, the recitation of 1 to 5 includes 1, 2, 3, 4, and 5, as well as fractions thereof, e.g., 1.5, 2.25, 3.75, 4.1, and the like) and any range within that range.

EXEMPLIFICATIONS

The following Examples have been included to provide guidance to one of ordinary skill in the art for practicing representative embodiments of the presently disclosed subject matter. In light of the present disclosure and the general level of skill in the art, those of skill can appreciate that the following Examples are intended to be exemplary only and that numerous changes, modifications, and alterations can be employed without departing from the scope of the presently disclosed subject matter. The synthetic descriptions and specific examples that follow are only intended for the purposes of illustration, and are not to be construed as limiting in any manner to make compounds of the disclosure by other methods.

EXAMPLE 1

Background

The development of CRISPR-Cas9 technology has revolutionized the field of gene-editing and offers a profoundly new approach to treating genetic diseases. The CRISPR-Cas9 system is composed of a guide RNA (gRNA) that targets the Cas9 nuclease in a sequence-specific fashion. Cleavage by the CRISPR system requires complementary base

pairing of the gRNA to a DNA sequence and the requisite protospaceradjacent motif (PAM), a short nucleotide motif found 3' to the target site (Dalkara, D. *et al. Science translational medicine* 5, 189ra176 (2013); Berns, KI *et al.* Fundamental Virology (ed B.N. Fields , and Knipe, D.M.) 545-562 (Raven Press, 1986).). Currently, the least restrictive and most commonly used Cas9 protein is from *S. pyogenes*, which recognizes the sequence NGG, and thus, the CRISPR targeting sequence is N20NGG. While numerous studies have shown that disease mutations can be efficiently targeted *in vitro*, the development of CRISPR-Cas9-based therapeutics for *in vivo* use is been hampered by safety concerns and delivery constraints.

While CRISPR targeting of disease mutations has been shown to be effective in numerous *in vitro* settings, and as well *in vivo* through mouse and other animal studies, all current approaches are still far from clinical use due in large part to delivery constraints. AAV vectors are the most frequently and successfully used
viral vectors in ocular gene therapy injection (Swiech, L. *et al. Nature biotechnology* 33, 102-106 (2015); Jinek, M. *et al. Science* 337, 816-821 (2012); Cong, L. *et al. Science* 339, 819-823 (2013); Mali, P. *et al. Science* 339, 823-826 (2013)). Several features make AAV the most attractive choice: the virus is nonpathogenic, it infects both dividing and non-dividing cells, expression can persist for long periods of time, and it is particularly noteworthy for its history of safety, efficacy and a general lack of toxicity in clinical trials. Additionally, specific AAV serotypes are effective in targeting photoreceptor cells after subretinal injection. While AAV vectors provide a safe means of delivering therapeutic CRISPR components, there is one major technical obstacle that limits their utility – their size. Wild type AAV genomes are ~4.7kb in length and recombinant viruses can package up to ~5.2kb (Mancuso, K. *et al. Nature* 461, 784-787 (2009); Beltran, WA *et al. Proceedings of the National Academy of Sciences of the United States of America* 109, 2132-2137 (2012)). This packaging capacity defines the upper limit of the DNA that can be used for a single viral vector.

The DNA required to express Cas9 and the gRNA, by conventional methods, exceeds 5.2kb: Pol II promoter (~0.5kb), SpCas9 (~4.1kb), Pol II terminator (~0.2kb), U6 promoter (~0.3kb), and the gRNA (~0.1kb). One approach to AAV delivery challenge is a two-vector approach: one AAV vector to deliver the Cas9, and another AAV vector for the gRNA (Petrs-Silva, H. *et al. Molecular therapy : the journal of the American Society of Gene Therapy* 19, 293-301 (2011)). However, the double AAV approach utilizes the small

mouse Mecp2 promoter, a gene that has been found to be expressed in retinal cells – with the critical exception of rods (Song, C. *et al. Epigenetics & chromatin* 7, 17 (2014); Jain, D. *et al. Pediatric neurology* 43, 35-40 (2010)) – suggesting that, aside from the potential toxicity due to increased viral delivery load, the co-delivery approach would likely fail to target the vast majority of LCA mutations a priori. While this is a potentially viable approach for other gene therapy-mediated genomic editing, provided herein is a single vector approach for retinal gene editing that should increase efficiency, target photoreceptors specifically, and reduce potential toxicity from viral load delivery.

It was recently reported that use of the H1 promoter, rather than the more traditionally used U6 promoter, to direct gRNA transcription allows an approximate doubling of the available CRISPR gene targeting space (Ranganathan, V *et al. Nature communications* 5, 4516 (2014)). Notably, it was also detected a lower propensity for off-target cutting suggesting that the H1 promoter would be more favorable for therapeutic approaches. During these studies, it was noticed the presence of a protein coding gene (PARP-2) in close genomic proximity to the endogenous H1RNA gene (Baer, M. *et al. Nucleic acids research* 18, 97-103 (1990); Myslinski, E. *et al. Nucleic acids research* 29, 2502-2509 (2001); Ame JC *et al. J Biol Chem.* 276(14):11092-9 (2001)). The sequence between the start of the H1RNA (a pol III RNA transcript) and the PARP-2 gene (a pol II transcript) is 230bp (FIG. 2), indicating that this relatively small sequence can function as a compact bidirectional promoter. It was hypothesized that these fortuitous properties of the H1 promoter might overcome the size hurdles of packaging both CRISPR components into a single AAV.

A. Enhancing H1 bidirectional pol II expression using 5'UTR modifications.

To develop H1 as a bidirectional pol II/III promoter, and because the poll III activity is already well characterized, an eGFP reporter construct was created to better optimize its pol II activity (FIG. 1A). Human (HEK293) and mouse cells (NIH3T3) demonstrated a weak, but clearly detectable GFP fluorescence, indicating that the H1 promoter could direct pol II expression. Using this GFP reporter system, experiments were performed to increase pol II expression while maintaining compactness by evaluating the three variable components in the system: (a) the promoter sequence, (b) the 5'UTR, and (c) the terminator sequence.

To evaluate different terminator sequences, seven different sequences were tested. It was found that the SV40 (240 bp and 120bp) terminators and a 49 bp synthetic poly(A)

sequence (SPA) were both functional for GFP expression, expression, although the SPA sequence did not work as well.

It was explored whether optimizing translation efficiency through modification of the 5'UTR would improve reporter expression. It was found that insertion of a 50 bp sequence taken from the beta-globin 5'UTR sequence was able to significantly improve reporter expression, and consistent with this notion, the simple insertion of 9 bases encoding a strong Kozak sequence (5'-GCCGCCACC-3' SEQ ID NO: 1) was sufficient to approximate these levels.

Also, using sequences identified as RNA sequences that mediate cap-independent initiation of translation (Wellensiek *et al. Nature Methods* 10, 747–750 (2013)) were also effective (6.947, and 6.967).

Focusing on the Kozak sequence, it was found that both a strong 9-base consensus (5'-GCCGCCACC-3' SEQ ID NO: 1) or a smaller 6-base consensus (5'-GCCACC-3' SEQ ID NO: 2) could enhanced GFP expression (FIG. 1). The 9-base sequence was slightly better consistently across different H1 bidirectional sequences.

B. Modulating bidirectional expression through use of different orthologous sequences.

Testing H1 promoter sequences from different organisms indicated that both mouse (176 bp) and rat (207 bp) sequences were able to drive stronger GFP expression than the human H1 promoter (~7 and ~6-fold higher, respectively). This result is contrary to that reported in other studies, for example Hung *et al.* (2006) *Biochemical and Biophysical Research Communication* 339:1035-1042, which reported that the human sequence was expressed at higher levels.

Genomic alignments were performed from the orthologous region of 36 eutherian mammals using the H1 promoter sequence. These provide additional sequences (SEQ ID NOs: 84-119) that can be used to fine-tune bidirectional expression.

C. Novel compact bidirectional promoter sequences with both pol II and pol III activity

After identifying that the H1 bidirectional promoter sequence could effectively direct the expression of both pol II and pol III genes, the genome for other potentially compact bidirectional sequences were searched.

A custom perl script was developed to compare the 5' transcriptional start sites of pol III genes with that of pol II genes. The results were filtered for those that are orientated

in opposite directions (divergent transcription). The input files could be annotated genome files or transcriptional data (pol II or pol III ChIP sites). Some ChIP sites that were identified from the human genome include, but not limited to:

ALOXE3-tRNA(lys):

TCTTTCCGCTCCAGGACCGCCCTGGGCCTGCAGGATCCTGGGCGGGAGC
CCAGGTGTCCGGGATCTGGGCCACTAGGGACTGGGGAGGAACCTCTCAGAGAA
GCCCATAGCCCGCAGCGGCCCCGCGCGGCCGGTTCCGGCGCCGCACTGTTCCA
GCCTCTACTATGGTACAGTCCCTGCGTCGCAGCCTCGGCGGGGGCTCTAAGAAC
GGGAGGCAGAAAAAGCTCAATCAGCAGCAGGCGAGCTTCACCCGCTGCTTCCA
AATCTGTGCCAAAATATTCTATGCTGCACAGATAAAATCCTCTGTCGGTTCTAC
AAGCCTGGCTTTTCCTATAGAGAACCCTCTTATAAGCAAAAAGTAAAGCTCTCG
TGAAGA;

C8orf41-SNORD13:

TCCTGACTGCAGCACCAGAAGGCTGGTCTCTCCCACAGAACGAGGATGG
AGGCGGGGAGGGATCCGTTGAAGAGGGAAGGAGCGATCACCCAAAGAGAACT
AAAATCAAATAAAATAAAACAGAGAGATGTCTTGGAGGAGGGGGCGAGTCTG
ACCGGGATAAGAATAAAGAGAAAGGGTGAACCCGGGAGGCGGAGTTTGCAGT
GAGCCGAGATCGCGCCACTGCACTCCAGCCTGGGCGACAGAGTGAGACTCCGT
CTCAGTAAAAAAAAAAAAAAAAAAAAGAATAAAGAGGAAAGGACGCAAGAA
AGGGAAAGGGGACTCTCAGGGAGTAAAAGAGTCTTACACTTTTAACAGTGACG
TTAAAAGACTACTGTTGCCTTTCTGAAGACTAAAAAGAAAAAAAACTTAAAAA
TTTAAAGAAATAAACTTCTGAGCCATGTCACCAACTTAACCACCCCCAGGTACC
TGCAACGGCTCGCGCCCGCCGGTGTCTAACAGGATCCGGACCTAGCTCATATTG
CTGCCGCAAAACGCAAGGCTAGCTTCCGCCAGTACTGCCGCAACACCTTCTTAT
TTCACGACGTATGGTCGTAAAGCAATAAAGATCCAGGCTCGGGAAAATGACGG
AGAGGTGGAACTATAGAGAATAAATTTGCATATATAATAATCCGCTCGCTAATT
GTGTTTCTGTTTTCCTTTGCTAAGGTAGAAACAAAAGAATAATCACAGAATCTC
AGTGGGACTTTGAAAATATCCAGGATTTTATACGTGAAGAATGGATGTATCGCA
TTACGGTAGTCACCCTATGTGTAAATTAGTGGCACATACTTGGCACTCCTTAAT
GTCAACTATAAGATG;

CGB1:

GTCTCTCTCTTAGCGGGATATCTTCCGCAAGCACTGGGAATGTGGACATG
GAAAGTAAATTGAGTCTCCGTGGGGGAGTGAGACAGGGAGTGAGGGGTGTTGG

ACGCGGCACGGGAACCTGGCCAGAGTCAGCGGACCCAATTGGCTGCTCTCTCT
CAGATGCAGTTCCCCTTCCTCCCTCCAGGGGGCGCCACGGAACGCAGGGCCCTC
ACTGGCCCTGGGGACTGGGTGACGTCAGGGATGAGCCTCTTGTGATTGGCTCCA
TCACCCTGCGTAAGATCAAAGGGAAGAAAGGATGGGCCCGACAA;

CGB2:

GTCTCTCTCTTAGCGGGATATCTTCCGCAAGCACTGGGGATGTGGACATG
GAAAGTAAATTGAGTCTCCGTGGGGGAGTGAGACAGGGAGTGAGGGGTGTTGG
ACGCGGCACGGGAACCCGGCCGGAGTCAGCGGACCCAATTGGCTGCTCTCTCT
CAGATACAGTTCCCCTTCCTCCCTCCAGGGGGCGCCACGGAACGCAGGGCCCTC
ACTGGCCCTGGGGACTGGGTGACGTCAGGGGTGAGCCTCTCCTGATTGGCTCCA
TCACCCTGCGTAAGGTCAAAAGGAAGAAAGGAGATCCCCGACAC;

DPP9-tRNA(gly):

TAACCGCTCAGCTGACCTCAGGAGGGCAGGGGTGCCTTCTAAAGGGTCC
AGAGAGCCTCCATTCCAGCTGCAGGCGTGGGACACAGACCGGGACGTGGGGCG
GCGGCCGGACTGGGCAGGTCGTCCCGGGTCCAGCGGCGCCTCACGGTCGCGGC
TCCATGCCCGGGACTGCGACCCCGGAAGTGGCGGGAGCGGGGGACGACAGCCG
CGGCGGACACAGGGGACCCGCCGGCTCAGGCACCTTTGACCCGGAAGTTGAGC
GACCCAGGCGGCGGCCTGGGATTGGACACCACCAGGCACGTACCAAGGCGTCC
GCGGCGCTTGGGGGGGAGCCCGCGGCGCGGCGGCCTAAGGTGCGTAACGCCCC
ATGAACGACATCTTCCGGTGGGTTAGGGAGAGACACCCCCCTGTGACTTGGTAT
CACTCAGTCAAACCCATGATCCCCCACTATTAAGGATATCCGGAGAGGATGCTA
CCTATCAGG;

MED16-RNU6-9:

GAGGGCAGTCACCAGCTCCTGGCCCGTGCGCCAAGCTCAGCGGGCGTCC
GCGGTGCGATCTTCCCTAGCGCCTCGGGTCTGGCGCCGCCATCTTCCTCGGTAA
CAACCAGTCGCCTGAGGCGTGGGGCCGCCTCCCAAAGACTTCTGGGAGGGCGG
TGCGGCTCAGGCTCTGCCCCGCCTCCGGGGCTATTTGCATACGACCAT

To validate this, the 7sk bidirectional promoter, or the region between the GSTA4 gene (pol II gene) and the RN7SK gene (pol III gene) were tested using our GFP reporter assay. The pol III activity of this promoter is well-documented.

Several variants of the 5' UTR were simultaneously tested; endogenously, the annotated 5'UTR for the human GSTA4 gene is split by an intron, which were fused when cloning into our reporter system (7sk1). A shortened form of the 5'UTR or the betaglobin

5'UTR was also tested (7sk3). All of the constructs used incorporated the strong 9-base Kozak consensus sequence (5'-GCCGCCACC-3' SEQ ID NO: 1) (7sk2).

The reporter assay indicated this region was able to direct pol II expression, confirming the ability to act as a bidirectional promoter region. Although this region is slightly larger than the H1 bidirectional promoter region, it was found to be more active.

The ability of orthologous sequences to direct expression was tested. The mouse sequence was also active, and slightly stronger than the human sequence.

These results confirmed the presence of bidirectional promoters with both pol II and pol III activity, and the ability to search genomic sequences for this regions.

D. An expression screen for bidirectional promoters with both RNA pol II and RNA pol III activity.

Bioinformatics:

Bidirectional promoter sequences could be used as starting points for the identification of different variants of pol II/pol III activity. A custom perl script was developed to compare the 5' transcriptional start sites of pol III genes with that of pol II genes. The results were filtered for those that are orientated in opposite directions (divergent transcription). The input files could be annotated genome files or transcriptional data (pol II or pol III ChIP sites). Using this information, bidirectional promoters were identified with both RNA pol II and RNA pol III activity. These sequences could be used to generated derivative sequences

Targeted transcription factor binding sites:

Promoter sequences could be used to identify transcription binding sites, or multiple promoter sequences could be aligned to identify transcription factor binding sites which could in turn be used to design a bidirectional promoter. Identification of transcription binding factor sites could be determined by consensus, or by using a differential distance matrix or multidimensional scaling (De Bleser P *et al. Genome Biol.* 2007;8(5):R83.).

Random sequence library:

Alternatively, a synthetic library of random sequences could be used as a starting point to screen for sequences with bidirectional activity. By randomizing, shuffling, or mutating the bidirectional sequence, one can search for DNA sequences that have different pol II or pol III promoter activity.

Screens could be set up using qPCR or fluorescence to read-out pol II and pol III activity. A simple fluorescence screen would use a reporter (GFP, mCherry, etc.) as

described, and an RNA aptamer, such as Broccoli that becomes fluorescent in the presence of a small molecule like DFHB1-1T. The screen could use protein readouts such as drug resistance (e.g. antibiotics), antigenic pepdies, or cell-surface markers, etc.

Alternative reporters could be function complexes that are comprised of both protein and RNA. For example, the MS2 coat protein binding stem loop could be incorporated into the RNA and the pol II transcription could encode the MS2 coat protein which would in turn bind the RNA sequence. If two binding sites are incorporated into the RNA, then fluorescence complementarity (split protein complementarity) could be used as a readout.

Screens can also be performed by separating and assessing the pol II activity and then the pol III activity. Finally, these sequences could then be combined in a fashion that allows for divergent transcription by both RNA pol II or RNA pol III.

Methods

Plasmid construction: The GFP reporter plasmids were constructed by Gibson Assembly (New England Biolabs, Ipswich, MA) (Gibson *et al.* (2009) *Nature* Methods 6:343-345) with slight modifications.

Human embryonic kidney (HEK) cell line 293T (Life Technologies, Grand Island, NY) was maintained at 37°C with 5% CO_2 / 20% O_2 in Dulbecco's modified Eagle's Medium (DMEM) (Invitrogen) supplemented with 10% fetal bovine serum (heat-inactivated) (Gibco, Life Technologies, Grand Island, NY) and 2mM GlutaMAX (Invitrogen).

N2A cell line (Life Technologies, Grand Island, NY) was maintained at 37°C with 5% CO_2 / 20% O_2 in Dulbecco's modified Eagle's Medium (DMEM) (Invitrogen) supplemented with 10% fetal bovine serum (Gibco, Life Technologies, Grand Island, NY) and 2mM GlutaMAX (Invitrogen).

Cells were transfected (Lipofectamine 3000) with each construct. GFP expression was analyzed using flow cytometry.

Human 7sk1 (GSTA4 5'UTR with Kozak):

GGTGGCGGCGATAGCTTTTCAGGCTTTCTGGAGTCCACTCGGAGGCCTGGAGCC
GCACAAAGCGCCAGGTCAGCGGTCCCGGCTGGGTGAGACCAGCAGGCGGCTCT
AGCGCGCGGGAGCTGGGCGAGGCTCCGGGACGACCTCACCAATGGAGACTGCA
GTATTTAGCATGCCCCACCCATCTGCAAGGCATTCTGGATAGTGTCAAAACAGC
CGGAAATCAAGTCCGTTTATCTCAAACTTTAGCATTTTGGGAATAAATGATATT

TGCTATGCTGGTTAAATTAGATTTTAGTTAAATTTCCTGCTGAAGCTCTAGTACG
ATAAGCAACTTGACCTAAGTGTAAAGTTGAGACTTCCTTCAGGTTTATATAGCT
TGTGCGCCGCTT GGGTACCTC (SEQ ID NO: 3)

Human 7sk2 (short 5' UTR):

GGTGGCGGCAGTATTTAGCATGCCCCACCCATCTGCAAGGCATTCTGGATAGTG
TCAAAACAGCCGGAAATCAAGTCCGTTTATCTCAAACTTTAGCATTTTGGGAAT
AAATGATATTTGCTATGCTGGTTAAATTAGATTTTAGTTAAATTTCCTGCTGAAG
CTCTAGTACGATAAGCAACTTGACCTAAGTGTAAAGTTGAGACTTCCTTCAGGT
TTATATAGCT TGTGCGCCGCTTGGGTACCTC (SEQ ID NO: 4)

Human 7sk3 (beta-globin 5' UTR):

GGTGTCTGTTTGAGGTTGCTAGTGAACACAGTTGTGTCAGAAGCAAATGTAGTA
TTTAGCATGCCCCACCCATCTGCAAGGCATTCTGGATAGTGTCAAAACAGCCGG
AAATCAAGTCCGTTTATCTCAAACTTTAGCATTTTGGGAATAAATGATATTTGC
TATGCTGGTTAAATTAGATTTTAGTTAAATTTCCTGCTGAAGCTCTAGTACGATA
AGCAACTTGACCTAAGTGTAAAGTTGAGACTTCCTTCAGGTTTATATAGCTTGT
GCGCCGCTTGG GTACCTC (SEQ ID NO: 5)

Mouse 7sk1 (GSTA4 5'UTR with Kozak):

GGTGGCGGCCTCGTATGGGACCGCACCGGACACGGGCGCCTGGGCCAGGAGCA
GAGCCGGGCCGTAGAATAGACATGGCCGTCGGGGGCGGGGCTTCGGAAGGTTT
AACCAATCCAAACTGTTGTATTTTGCATAGCCCCAAAGCATTTTGGTTAACAGT
AAAAACATCCTAAATTTAAGTATTTTAATTTAAACTTAGAACGAAGCGAGTATA
AAAAGGATTATTTAACCCTAAAACGGATTCAGGATTTGTTATAATATCAAGTAC
AGTCGGCTACATAAGGTCACCACATGTGTAAAGTTACAAAATTCTATGGCCTTA
TATACCTACCAAGA GCCTGCTTACTCTC (SEQ ID NO: 6)

Mouse 7sk2 (short 5' UTR):

GGTGGCGGCAGACATGGCCGTCGGGGGCGGGGCTTCGGAAGGTTTAACCAATC
CAAACTGTTGTATTTTGCATAGCCCCAAAGCATTTTGGTTAACAGTAAAAACAT
CCTAAATTTAAGTATTTTAATTTAAACTTAGAACGAAGCGAGTATAAAAAGGAT
TATTTAACCCTAAAACGGATTCAGGATTTGTTATAATATCAAGTACAGTCGGCT
ACATAAGGTCACCACATGTGTAAAGTTACAAAATTCTATGGCCTTATATACCTA
CCAAGAGCCTGCTT ACTCTC (SEQ ID NO: 7)

Mouse 7sk3 (beta-globin 5' UTR):

GGTGTCTGTTTGAGGTTGCTAGTGAACACAGTTGTGTCAGAAGCAAATGTAGAC

ATGGCCGTCGGGGGCGGGGCTTCGGAAGGTTTAACCAATCCAAACTGTTGTATT
TTGCATAGCCCCAAAGCATTTTGGTTAACAGTAAAAACATCCTAAATTTAAGTA
TTTTAATTTAAACTTAGAACGAAGCGAGTATAAAAAGGATTATTTAACCCTAAA
ACGGATTCAGGATTTGTTATAATATCAAGTACAGTCGGCTACATAAGGTCACCA
CATGTGTAAAGTTACAAAATTCTATGGCCTTATATACCTACCAAGAGCCTGCTT
ACTCTC (SEQ ID NO: 8)

1. Dryja, TP *et al. The New England Journal of Medicine* 323, 1302-1307 (1990).
2. Dryja, TP *et al. Nature* 343, 364-366 (1990).
3. Doudna, JA *et al. Science* 346, 1258096 (2014).
4. Hsu, PD *et al. Cell* 157, 1262-1278 (2014).
5. Liang, Y. *et al. The Journal of Biological Chemistry* 279, 48189-48196 (2004).
6. Dalkara, D. & Sahel, J. A. *Comptes Rendus Biologies* 337, 185-192 (2014).
7. Day, TP. *et al. Advances in Experimental Medicine and Biology* 801, 687-693 (2014).
8. Willett, K. & Bennett, Frontiers in immunology 4, 261, (2013).
9. Dinculescu, A. *et al. Human Gene Therapy* 16, 649-663 (2005).
10. Kotterman, MA *et al. Gene Therapy* 22, 116-126 (2015).
11. Mowat, FM *et al. Gene Therapy* 21, 96-105 (2014).
12. Dalkara, D. *et al. Science Translational Medicine* 5, 189ra176 (2013).
13. Berns, KI *et al.* Fundamental Virology (ed B.N. Fields , and Knipe, D.M.) 545-562 (Raven Press, 1986).
14. Swiech, L. *et al. Nature Biotechnology* 33, 102-106 (2015).
15. Jinek, M. *et al. Science* 337, 816-821 (2012).
16. Cong, L. *et al. Science* 339, 819-823 (2013).
17. Mali, P. *et al. Science* 339, 823-826 (2013).
18. Mancuso, K. *et al. Nature* 461, 784-787 (2009).
19. Beltran, WA *et al. Proceedings of the National Academy of Sciences of the United States of America* 109, 2132-2137 (2012).
20. Petrs-Silva, H. *et al. Molecular therapy : The Journal of the American Society of Gene Therapy* 19, 293-301 (2011).
21. Song, C. *et al. Epigenetics & Chromatin* 7, 17 (2014).
22. Jain, D. *et al. Pediatric Neurology* 43, 35-40 (2010).
23. Ranganathan, V *et al. Nature Communications* 5, 4516 (2014).

24. Baer, M. *et al. Nucleic Acids Research* 18, 97-103 (1990).

25. Myslinski, E. *et al. Nucleic Acids Research* 29, 2502-2509 (2001).

26. Ame JC *et al. J Biol Chem*. 276(14):11092-9 (2001).

EXAMPLE 2

Conditional pol II/pol III bidirectional promoter expression for regulating ribonucleoprotein enzymatic activity or RNA-directed nucleases

Previous work has demonstrated that the pol III activity of the H1 promoter can engineered into a tet-responsive promoter and can be regulated by tetracycline. In the presence of the Tet repressor (TetR), the Tet operator (TetO) sequences engineered into the H1 promoter are bound by TetR and repress pol III expression.

This system has not been previously combined with the bidirectional component of the H1 promoter, a system that could allow for regulation of pol II and pol III transcripts. Ribonucleoprotein enzymes or RNA-directed nucleases, could be regulated by tetracycline using an engineered bidirectional promoter system.

Here, it was shown that the pol II activity of H1 was not drastically altered by the presence of TetO operator sites. This provides a mechanism for the inducible activity of ribonucleoprotein enzymes or RNA-directed nucleases, such as Cas9/gRNA. Additionally, placement of the TetO sites could be placed as to specifically repress either the pol II or pol III transcripts, or both.

Such a system would have tremendous clinical advantages for regulating CRISPR activity delivered by AAV viruses and could be used to other compact bidirectional promoters.

1. Matthess *et al. Oncogene*. 2005 Apr 21;24(18):2973-80.

Methods

Plasmid construction: The GFP reporter plasmids were constructed by Gibson Assembly (New England Biolabs, Ipswich, MA) (Gibson *et al.* (2009) *Nature* Methods 6:343-345) with slight modifications.

Human embryonic kidney (HEK) cell line 293T (Life Technologies, Grand Island, NY) was maintained at 37°C with 5% CO_2 / 20% O_2 in Dulbecco's modified Eagle's Medium (DMEM) (Invitrogen) supplemented with 10% fetal bovine serum (heat-inactivated) (Gibco, Life Technologies, Grand Island, NY) and 2mM GlutaMAX (Invitrogen).

N2A cell line (Life Technologies, Grand Island, NY) was maintained at 37°C with 5% CO_2 / 20% O_2 in Dulbecco's modified Eagle's Medium (DMEM) (Invitrogen) supplemented with 10% fetal bovine serum (Gibco, Life Technologies, Grand Island, NY) and 2mM GlutaMAX (Invitrogen).

Cells were transfected (Lipofectamine 3000) with each construct. GFP expression was analyzed using flow cytometry.

mm079 target sequence: GAAGAAGGTTCGAGATCTCA (SEQ ID NO: 9)

mm079 genomic target site: GAAGAAGGTTCGAGATCTCAAGG (SEQ ID NO: 10)

TetO site: TCCCTATCAGTGATAGAGA (SEQ ID NO: 11)

H1_wt:

GGAATTCGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAGTG TCACTAGGCGGGAACACCCAGCGCGCGTGCGCCCTGGCAGGAAGATGGCTGTG AGGGACAGGGGAGTGGCGCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGA AATCACCATAAACGTGAAATGTCTTTGGATTTGGGAATCTTATAAGTTCTGTAT GAGACCACTTTTTCCC (SEQ ID NO: 12)

H1_TetO:

GGAATTCGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAGTG TCACTAGGCGGGAACACCCAGCGCGCGTGCGCCCTGGCAGGAAGATGGCTGTG AGGGACAGGGGAGTGGCGCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGA AATCACCATAAACGTGAAATCCCTATCAGTGATAGAGACTTATAAGTTCCCTAT CAGTGATAGAGATCCC (SEQ ID NO: 13)

TetR sequence:

MSRLDKSKVINSALELLNEVGIEGLTTRKLAQKLGVEQPTLYWHVKNKRALLDAL AIEMLDRHHTHFCPLEGESWQDFLRNNAKSFRCALLSHRDGAKVHLGTRPTEKQY ETLENQLAFLCQQGFSLENALYALSAVGHFTLGCVLEDQEHQVAKEERETPTTDS MPPLLRQAIELFDHQGAEPAFLFGLELIICGLEKQLKCESGSAYSGSREF (SEQ ID NO: 14)

EXAMPLE 3

A. Regulation of Genome-Engineering Nucleases through Post-Transcriptional Cell-cycle regulation

The advent of new RNA-directed nucleases has revolutionized genome-editing technology and transformed biological research. These technologies, while providing an unprecedented means to modify genomes with precision, are constrained by cellular

processes that modulate how DNA breaks are repaired. For genome-engineering applications, in particular, the ability to modulate a cells response to DNA breaks has enormous consequences. (see Gutschner *et. al.* (2016) Cell Reports).

Cellular response to DNA breaks occurs primarily through one of two competing pathways: nonhomologous end-joining (NHEJ), or Homology Directed Repair (HDR). NHEJ is generally considered to be an error-prone pathway which results sequence changes around the break point. In general, NHEJ is the more efficient or dominant pathway for DNA repair. On the other hand, HDR pathways are far less error-prone, but require stretches of homology to template and repair.

Both pathways have properties that are favorable for different outcomes. For example, if one wishes to "knock-out" a gene, NHEJ is the preferable pathway, as DNA breaks in that gene will largely result in sequence changes. In the laboratory, many genetic screens are dependent on eliminating gene function, and for therapeutic approaches, complete disruption of a gene with a dominant or gain-of-function mutation would be highly desirable. However, if one wishes to introduce a specific sequence change ("knock-in"), one would try to favor repair though a HDR pathway thus allowing for precise cut and paste outcomes. For many scientific applications and most clinical applications this pathway is favorable, as it can allow for precise mutation repair.

It is also known that cellular response to DNA breaks is modulated by the cell-cycle (FIG. 9A); the phase of the cell-cycle largely dictates the choice of DNA pathway (FIG. 9B). NHEJ predominates through G1 of the cell-cycle, while HDR predominates during S and G2 phases. Thus, the ability to modulate the timing of DNA breaks with respect to the cell-cycle, can be a powerful method to bias the cell towards a specific repair pathway.

Many proteins are regulated in cell-cycle-dependent fashion, largely through transcriptional regulation and post-transcriptional mechanisms, notably ubiquitin-mediated proteolysis. Generally, regulation through transcriptional mechanisms are slower, while proteolytic mechanisms are rapid. For genome-editing applications, proteolytic mechanisms likely offer a more favorable method of regulation.

Two well-characterized proteins with opposing, and thus oscillating, expression in the cell are Geminin (Gem) and Cdt1. Cdt1 accumulates in the G1 phase of the cell-cycle, while Gem accumulates during S/G2/M phases. The specific regions of these proteins that are required for cell-cycle regulation have been mapped: amino acids 30-120 for hCdt1, and amino acids 1-110 for hGem (Sakaue-Sawano A *et al. Cell* 132, 487-498 (2008)).

Furthermore, proteins fused to these domains can be made to exhibit cell-cycle-dependent regulation, even if the respective mRNA is constantly being transcribed throughout the cell-cycle. In essence, cellular proteins only recognize the domain during specific phases of the cell-cycle, which results in ubiquitination and then rapid degradation of the fusion protein.

Here, it was shown that genome-editing technologies, such as RNA-directed nucleases, can be fused to either hGem of hCdt1 domains, resulting in cell-cycle dependent regulation of these proteins (FIG. 9C). Consequently, this regulation results in drastically different DNA repair outcomes. Using the CRISPR-Cas9 system, it was shown that either NHEJ and HDR pathways were enhanced or suppressed using Cas9 fused to either hGem or hCdt1. The system is so robust that it can increase the level of HDR to levels greater than NHEJ, which is highly significant considering that HDR is far less efficient in the cell.

Furthermore, the method described is tremendously powerful in its elegance and simplicity. Other attempts to regulate the outcome of genome-editing technologies, have utilized cumbersome methods of arresting cells in certain phases of the cell-cycle followed by technically challenging and inefficient methods of direct protein/RNA delivery. Instead, the method described here relies on simple delivery of a plasmid, mRNA, or protein encoding a fusion to the respective regulatory sequences.

Methods

Plasmid construction: To generate the cell-cycle regulated nucleases proteins, the human codon optimized Cas9 gene from a Cas9:T2A:GFP vector (Addgene 44719) was modified to replace the T2A with a better cleaving P2A sequence. Next the GFP fluorescent reporter was replace with the sequence encoding the mKate2 fluorescent protein. Next, the vector was linearized and gBlocks encoding a flexible 15 amino acid sequence linker fused to either the hGem(1-110) or hCdt1(30-120) domains were inserted in frame to generate Cas9:linker:hGem and Cas9:linker:hCdt1 (referred to as Cas9-Gem or Cas9-Cdt, respectively). All cloning steps were preformed using Gibson Assembly (New England Biolabs, Ipswich, MA) (Gibson *et al.* (2009) *Nature Methods* 6:343-345) with slight modifications.

Human embryonic kidney (HEK) cell line 293T (Life Technologies, Grand Island, NY) was maintained at 37°C with 5% CO_2 / 20% O_2 in Dulbecco's modified Eagle's Medium (DMEM) (Invitrogen) supplemented with 10% fetal bovine serum (Gibco, Life Technologies, Grand Island, NY) and 2mM GlutaMAX (Invitrogen).

A gRNA targeting Rhodopsin (hs086172175: AGTACTGTGGGTACTCGAAGGGG (SEQ ID NO: 68)) (see Jaskula-Ranga, V., & Zack, D. J. (2016). grID: A CRISPR-Cas9 guide RNA Database and Resource for Genome-Editing. *bioRxiv*, 097352) was generated by overlapping oligos that were annealed and amplified by PCR using two-step amplification Phusion Flash DNA polymerase (Thermo Fisher Scientific, Rockford, IL), and subsequently purified using Zymo DNA clean and concentrator columns. The purified PCR products were then resuspended in H2O and quantitated using a NanoDrop 1000 (Thermo Fisher Scientific). The gRNA-expressing constructs were generated using the Gibson Assembly (New England Biolabs, Ipswich, MA) (Gibson *et al. Nature Methods* 6:343-345 (2009)) with slight modifications. The total reaction volume was reduced from 20µl to 2µl.

HEK293 cells were co-transfected with Cas9 (unmodified, or cell-cycle regulated versions) and the gRNA construct targeting rhodopsin. 48hrs post transfection, genomic DNA was harvested and the sequence surrounding the target cut sites were amplified according to the primers listed below. The PCR products were then purified and quantitated before performing the T7 Endo I assay. Briefly, 200ng of PCR product was denatured and then slowly re-annealed to allow for the formation of heteroduplexes, T7 Endonuclease I was added to the PCR products and incubated at 37C for 25 minutes to cleave heteroduplexes, the reaction was quenched in loading dye, and finally, the reaction was run on a 6% TBE PAGE gel to resolve the products. The gel was stained with SYBR-Gold, visualized, and quantitated using ImageJ. NHEJ frequencies were calculated using the binomial-derived equation:

$$1-\sqrt{1-\frac{(a+b)}{(a+b+c)}}\ \mathrm{X}\ 100$$

; where the values of "a" and "b" are equal to the integrated area of the cleaved fragments after background subtraction and "c" is equal to the integrated area of the uncleaved PCR product after background subtraction. To calculate HDR frequencies, the same PCR product as above, prior to heteroduplex formation, was mixed with EcoRI for 1hr, then quenched in loading dye, and finally, the reaction was run on a 6% TBE PAGE gel to resolve the products.

HDR frequencies were calculated using the equation:

$$\%\ \mathrm{HDR}=\frac{(a+b)}{(a+b+c)}*100$$

; where the values of "a" and "b" are equal to the integrated area of the cleaved fragments after background subtraction and "c" is equal to the integrated area of the un-cleaved PCR product after background subtraction.

HDR oligo (EcoRI):

CAGAAGGCCCTAACTTCTACGTGCCCTTCTCCAATGCGACGGGTGTGGTACGCA GCCCCTGAATTCACCCACAGTACTACCTGGCTGAGCCATGGCAGTTCTCCATGC TGGCCGCCTACATGTTTC (SEQ ID NO: 15)

T7EI primers:

Rho_HDR_F; TGGAGCCCTGAGTGGCTGAG (SEQ ID NO: 16)

Rho_HDR_R; CCACCTAGGACCATGAAGAGGTCAG (SEQ ID NO: 17)

Protein Sequences:

hCdt1(30-120)

PSPARPALRAPASATSGSRKRARPPAAPGRDQARPPARRRLRLSVDEVSSPSTPEAP DIPACPSPGQKIKKSTPAAGQPPHLTSAQDQDTI (SEQ ID NO: 18)

hGem (1-110)

MNPSMKQKQEEIKENIKNSSVPRRTLKMIQPSASGSLVGRENELSAGLSKRKHRND HLTSTTSSPGVIVPESSENKNLGGVTQESFDLMIKENPSSQYWKEVAEKRRKAL (SEQ ID NO: 19)

Cas9:linker:hCdt1

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSG ETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFL IEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIA QLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIG DQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLA RGNSRFAWMTRKSEETITPWNFEE VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLG TYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK EDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVI

EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI
TKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF
VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEK
NPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGDSRADPKKKRKVRGGGGSGGGGSGGGGSPSPARPA
LRAPASATSGSRKRARPPAAPGRDQARPPARRRLRLSVDEVSSPSTPEAPDIPACPSP
GQKIKKSTPAAGQPPHLTSAQDQDTI (SEQ ID NO: 20)

Cas9:linker:hGem

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSG
ETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFL
IEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIA
QLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIG
DQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED
LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLA
RGNSRFAWMTRKSEETITPWNFEE
VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMR
KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLG
TYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK
QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK
EDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVI
EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI
TKHVAQILDSRMNTKYDENDKLIRE

VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF
VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEK
NPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGDSRADPKKKRKVRGGGGSGGGGSGGGGSMNPSMK
QKQEEIKENIKNSSVPRRTLKMIQPSASGSLVGRENELSAGLSKRKHRNDHLTSTTS
SPGVIVPESSENKNLGGVTQESFDLMIKENPSSQYWKEVAEKRRKAL (SEQ ID NO: 21)

Cdt1:linker:Cas9

MPSPARPALRAPASATSGSRKRARPPAAPGRDQARPPARRRLRLSVDEVSSPSTPEA
PDIPACPSPGQKIKKSTPAAGQPPHLTSAQDQDTIGGGGSGGGGSGGGGSDKKYSIG
LDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGN
IVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK
NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLF
LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT
FDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFA
WMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIEC
FDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMI
EERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG
FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV
VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEH
PVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDN
KVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGL
SELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSD
FRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVR
KMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG

RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG
GFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKL
KGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGDSRADPKKKRKVR (SEQ ID NO: 22)

Gem:linker:Cas9

MNPSMKQKQEEIKENIKNSSVPRRTLKMIQPSASGSLVGRENELSAGLSKRKHRND
HLTSTTSSPGVIVPESSENKNLGGVTQESFDLMIKENPSSQYWKEVAEKRRKALGG
GGSGGGGSGGGGSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIK
KNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRL
EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL
AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSAR
LSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYD
DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHH
QDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG
TEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKI
LTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFD
KNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK
TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE
NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLI
NGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHI
ANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRE
RMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYW
RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRM
NTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG
TALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSK
ESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKE
LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGEL
QKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK

RYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKVR (SEQ ID NO: 23)

Cas9:Cdt1:

ATGGACAAGAAGTACTCCATTGGGCTCGATATCGGCACAAACAGCGTCGGCTG
GGCCGTCATTACGGACGAGTACAAGGTGCCGAGCAAAAAATTCAAAGTTCTGG
GCAATACCGATCGCCACAGCATAAAGAAGAACCTCATTGGCGCCCTCCTGTTCG
ACTCCGGGGAGACGGCCGAAGCCACGCGGCTCAAAAGAACAGCACGGCGCAG
ATATACCCGCAGAAAGAATCGGATCTGCTACCTGCAGGAGATCTTTAGTAATG
AGATGGCTAAGGTGGATGACTCTTTCTTCCATAGGCTGGAGGAGTCCTTTTTGG
TGGAGGAGGATAAAAAGCACGAGCGCCACCCAATCTTTGGCAATATCGTGGAC
GAGGTGGCGTACCATGAAAAGTACCCAACCATATATCATCTGAGGAAGAAGCT
TGTAGACAGTACTGATAAGGCTGACTTGCGGTTGATCTATCTCGCGCTGGCGCA
TATGATCAAATTTCGGGGACACTTCCTCATCGAGGGGGACCTGAACCCAGACA
ACAGCGATGTCGACAAACTCTTTATCCAACTGGTTCAGACTTACAATCAGCTTT
TCGAAGAGAACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAATCCTGAGC
GCTAGGCTGTCCAAATCCCGGCGGCTCGAAAACCTCATCGCACAGCTCCCTGGG
GAGAAGAAGAACGGCCTGTTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACC
CCCAACTTTAAATCTAACTTCGACCTGGCCGAAGATGCCAAGCTTCAACTGAGC
AAAGACACCTACGATGATGATCTCGACAATCTGCTGGCCCAGATCGGCGACCA
GTACGCAGACCTTTTTTTGGCGGCAAAGAACCTGTCAGACGCCATTCTGC
TGAGTGATATTCTGCGAGTGAACACGGAGATCACCAAAGCTCCGCTGAGCGCT
AGTATGATCAAGCGCTATGATGAGCACCACCAAGACTTGACTTTGCTGAAGGC
CCTTGTCAGACAGCAACTGCCTGAGAAGTACAAGGAAATTTTCTTCGATCAGTC
TAAAAATGGCTACGCCGGATACATTGACGGCGGAGCAAGCCAGGAGGAATTTT
ACAAATTTATTAAGCCCATCTTGGAAAAAATGGACGGCACCGAGGAGCTGCTG
GTAAAGCTTAACAGAGAAGATCTGTTGCGCAAACAGCGCACTTTCGACAATGG
AAGCATCCCCCACCAGATTCACCTGGGCGAACTGCACGCTATCCTCAGGCGGC
AAGAGGATTTCTACCCCTTTTTGAAAGATAACAGGGAAAAGATTGAGAAAATC
CTCACATTTCGGATACCCTACTATGTAGGCCCCCTCGCCCGGGGAAATTCCAGA
TTCGCGTGGATGACTCGCAAATCAGAAGAGACCATCACTCCCTGGAACTTCGA
GGAAGTCGTGGATAAGGGGGCCTCTGCCCAGTCCTTCATCGAAAGGATGACTA
ACTTTGATAAAAATCTGCCTAACGAAAAGGTGCTTCCTAAACACTCTCTGCTGT
ACGAGTACTTCACAGTTTATAACGAGCTCACCAAGGTCAAATACGTCACAGAA

GGGATGAGAAAGCCAGCATTCCTGTCTGGAGAGCAGAAGAAAGCTATCGTGGA
CCTCCTCTTCAAGACGAACCGGAAAGTTACCGTGAAACAGCTCAAAGAAGACT
ATTTCAAAAAGATTGAATGTTTCGACTCTGTTGAAATCAGCGGAGTGGAGGATC
GCTTCAACGCATCCCTGGGAACGTATCACGATCTCCTGAAAATCATTAAA
GACAAGGACTTCCTGGACAATGAGGAGAACGAGGACATTCTTGAGGACATTGT
CCTCACCCTTACGTTGTTTGAAGATAGGGAGATGATTGAAGAACGCTTGAAAAC
TTACGCTCATCTCTTCGACGACAAAGTCATGAAACAGCTCAAGAGGCGCCGAT
ATACAGGATGGGGGCGGCTGTCAAGAAAACTGATCAATGGGATCCGAGACAAG
CAGAGTGGAAAGACAATCCTGGATTTTCTTAAGTCCGATGGATTTGCCAACCGG
AACTTCATGCAGTTGATCCATGATGACTCTCTCACCTTTAAGGAGGACATCCAG
AAAGCACAAGTTTCTGGCCAGGGGGACAGTCTTCACGAGCACATCGCTAATCTT
GCAGGTAGCCCAGCTATCAAAAAGGGAATACTGCAGACCGTTAAGGTCGTGGA
TGAACTCGTCAAAGTAATGGGA
AGGCATAAGCCCGAGAATATCGTTATCGAGATGGCCCGAGAGAACCAAACTAC
CCAGAAGGGACAGAAGAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGT
ATAAAAGAACTGGGGTCCCAAATCCTTAAGGAACACCCAGTTGAAAACACCCA
GCTTCAGAATGAGAAGCTCTACCTGTACTACCTGCAGAACGGCAGGGACATGT
ACGTGGATCAGGAACTGGACATCAATCGGCTCTCCGACTACGACGTGGATCAT
ATCGTGCCCCAGTCTTTTCTCAAAGATGATTCTATTGATAATAAAGTGTTGACA
AGATCCGATAAAAATAGAGGGAAGAGTGATAACGTCCCCTCAGAAGAAGTTGT
CAAGAAAATGAAAAATTATTGGCGGCAGCTGCTGAACGCCAAACTGATCACAC
AACGGAAGTTCGATAATCTGACTAAGGCTGAACGAGGTGGCCTGTCTGAGTTG
GATAAAGCCGGCTTCATCAAAAGGCAGCTTGTTGAGACACGCCAGATCACCAA
GCACGTGGCCCAAATTCTCGATTCACGCATGAACACCAAGTACGATGAAAATG
ACAAACTGATTCGAGAGGTGAAAGTTATTACTCTGAAGTCTAAGCTGGTCTCAG
ATTTCAGAAAGGACTTTCAGTTTTATAAGGTGAGAGAGATCAACAATTACCACC
ATGCGCATGATGCCTACCTGAATGCAGTGGTAGGCACTGCACTTATCAAAAAAT
ATCCCAAGCTTGAATCTGAATTTGTTTACGGAGACTATAAAGTGTACGATGTTA
GGAAAATGATCGCAAAGTCTGAGCAGGAAATAGGCAAGGCCACCGCTAAGTAC
TTCTTTTACAGCAATATTATGAATTTTTTCAAGACCGAGATTACACTGGCCAAT
GGAGAGATTCGGAAGCGACCACTTATCGAAACAAACGGAGAAACAGGAGAAA
TCGTGTGGGACAAGGGTAGGGATTTCGCGACAGTCCGGAAGGTCCTGTCCATG
CCGCAGGTGAACATCGTTAAAAAGACCGAAGTACAGACCGGAGGCTTCTCCAA

GGAAAGTATCCTCCCGAAAAGGAACAGCGACAAGCTGATCGCACGCAAAAAA
GATTGGGACCCCAAGAAATACGGCGGATTCGATTCTCCTACAGTCGCTTACAGT
GTACTGGTTGTGGCCAAAGTGGAGAAAGGGAAGTCTAAAAAACTCAAAAGCGT
CAAGGAACTGCTGGGCATCACAATCATGGAGCGATCAAGCTTCGAAAAAAACC
CCATCGACTTTCTCGAGG
CGAAAGGATATAAAGAGGTCAAAAAAGACCTCATCATTAAGCTTCCCAAGTAC
TCTCTCTTTGAGCTTGAAAACGGCCGGAAACGAATGCTCGCTAGTGCGGGCGA
GCTGCAGAAAGGTAACGAGCTGGCACTGCCCTCTAAATACGTTAATTTCTTGTA
TCTGGCCAGCCACTATGAAAAGCTCAAAGGGTCTCCCGAAGATAATGAGCAGA
AGCAGCTGTTCGTGGAACAACACAAACACTACCTTGATGAGATCATCGAGCAA
ATAAGCGAATTCTCCAAAAGAGTGATCCTCGCCGACGCTAACCTCGATAAGGT
GCTTTCTGCTTACAATAAGCACAGGGATAAGCCCATCAGGGAGCAGGCAGAAA
ACATTATCCACTTGTTTACTCTGACCAACTTGGGCGCGCCTGCAGCCTTCAAGT
ACTTCGACACCACCATAGACAGAA
AGCGGTACACCTCTACAAAGGAGGTCCTGGACGCCACACTGATTCATCAGTCA
ATTACGGGGCTCTATGAAACAAGAATCGACCTCTCTCAGCTCGGTGGAGACAG
CAGGGCTGACCCCAAGAAGAAGAGGAAGGTGAGGGGAGGCGGAGGATCTGGC
GGAGGCGGAAGTGGCGGAGGGGGCAGCCCATCTCCTGCCAGACCTGCTCTGAG
AGCCCCTGCCTCTGCCACAAGCGGCAGCAGAAAGAGAGCCAGACCTCCTGCCG
CCCCTGGCAGAGATCAGGCTAGACCTCCAGCTCGGCGGAGACTGAGACTGAGC
GTGGACGAGGTGTCCAGCCCTAGCACACCTGAGGCCCCTGATATCCCCGCCTGT
CCTAGCCCTGGCCAGAAGATCAAGAAGTCCACCCCTGCCGCCGGACAGCCTCC
TCATCTGACATCTGCCCAGGACCAGGACACCATC (SEQ ID NO: 24)

Cas9:Gem:

ATGGACAAGAAGTACTCCATTGGGCTCGATATCGGCACAAACAGCGTCGGCTG
GGCCGTCATTACGGACGAGTACAAGGTGCCGAGCAAAAAATTCAAAGTTCTGG
GCAATACCGATCGCCACAGCATAAAGAAGAACCTCATTGGCGCCCTCCTGTTCG
ACTCCGGGGAGACGGCCGAAGCCACGCGGCTCAAAAGAACAGCACGGCGCAG
ATATACCCGCAGAAAGAATCGGATCTGCTACCTGCAGGAGATCTTTAGTAATG
AGATGGCTAAGGTGGATGACTCTTTCTTCCATAGGCTGGAGGAGTCCTTTTTGG
TGGAGGAGGATAAAAAGCACGAGCGCCACCCAATCTTTGGCAATATCGTGGAC
GAGGTGGCGTACCATGAAAAGTACCCAACCATATATCATCTGAGGAAGAAGCT
TGTAGACAGTACTGATAAGGCTGACTTGCGGTTGATCTATCTCGCGCTGGCGCA

TATGATCAAATTTCGGGGACACTTCCTCATCGAGGGGGACCTGAACCCAGACA
ACAGCGATGTCGACAAACTCTTTATCCAACTGGTTCAGACTTACAATCAGCTTT
TCGAAGAGAACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAATCCTGAGC
GCTAGGCTGTCCAAATCCCGGCGGCTCGAAAACCTCATCGCACAGCTCCCTGGG
GAGAAGAAGAACGGCCTGTTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACC
CCCAACTTTAAATCTAACTTCGACCTGGCCGAAGATGCCAAGCTTCAACTGAGC
AAAGACACCTACGATGATGATCTCGACAATCTGCTGGCCCAGATCGGCGACCA
GTACGCAGACCTTTTTTTGGCGGCAAAGAACCTGTCAGACGCCATTCTGC
TGAGTGATATTCTGCGAGTGAACACGGAGATCACCAAAGCTCCGCTGAGCGCT
AGTATGATCAAGCGCTATGATGAGCACCACCAAGACTTGACTTTGCTGAAGGC
CCTTGTCAGACAGCAACTGCCTGAGAAGTACAAGGAAATTTTCTTCGATCAGTC
TAAAAATGGCTACGCCGGATACATTGACGGCGGAGCAAGCCAGGAGGAATTTT
ACAAATTTATTAAGCCCATCTTGGAAAAAATGGACGGCACCGAGGAGCTGCTG
GTAAAGCTTAACAGAGAAGATCTGTTGCGCAAACAGCGCACTTTCGACAATGG
AAGCATCCCCCACCAGATTCACCTGGGCGAACTGCACGCTATCCTCAGGCGGC
AAGAGGATTTCTACCCCTTTTTGAAAGATAACAGGGAAAAGATTGAGAAAATC
CTCACATTTCGGATACCCTACTATG
TAGGCCCCCTCGCCCGGGGAAATTCCAGATTCGCGTGGATGACTCGCAAATCA
GAAGAGACCATCACTCCCTGGAACTTCGAGGAAGTCGTGGATAAGGGGGCCTC
TGCCCAGTCCTTCATCGAAAGGATGACTAACTTTGATAAAAATCTGCCTAACGA
AAAGGTGCTTCCTAAACACTCTCTGCTGTACGAGTACTTCACAGTTTATAACGA
GCTCACCAAGGTCAAATACGTCACAGAAGGGATGAGAAAGCCAGCATTCCTGT
CTGGAGAGCAGAAGAAAGCTATCGTGGACCTCCTCTTCAAGACGAACCGGAAA
GTTACCGTGAAACAGCTCAAAGAAGACTATTTCAAAAAGATTGAATGTTTCGA
CTCTGTTGAAATCAGCGGAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGT
ATCACGATCTCCTGAAAATCATTAAA
GACAAGGACTTCCTGGACAATGAGGAGAACGAGGACATTCTTGAGGACATTGT
CCTCACCCTTACGTTGTTTGAAGATAGGGAGATGATTGAAGAACGCTTGAAAAC
TTACGCTCATCTCTTCGACGACAAAGTCATGAAACAGCTCAAGAGGCGCCGAT
ATACAGGATGGGGGCGGCTGTCAAGAAAACTGATCAATGGGATCCGAGACAAG
CAGAGTGGAAAGACAATCCTGGATTTTCTTAAGTCCGATGGATTTGCCAACCGG
AACTTCATGCAGTTGATCCATGATGACTCTCTCACCTTTAAGGAGGACATCCAG
AAAGCACAAGTTTCTGGCCAGGGGGACAGTCTTCACGAGCACATCGCTAATCTT

GCAGGTAGCCCAGCTATCAAAAAGGGAATACTGCAGACCGTTAAGGTCGTGGA
TGAACTCGTCAAAGTAATGGGA
AGGCATAAGCCCGAGAATATCGTTATCGAGATGGCCCGAGAGAACCAAACTAC
CCAGAAGGGACAGAAGAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGT
ATAAAAGAACTGGGGTCCCAAATCCTTAAGGAACACCCAGTTGAAAACACCCA
GCTTCAGAATGAGAAGCTCTACCTGTACTACCTGCAGAACGGCAGGGACATGT
ACGTGGATCAGGAACTGGACATCAATCGGCTCTCCGACTACGACGTGGATCAT
ATCGTGCCCCAGTCTTTTCTCAAAGATGATTCTATTGATAATAAAGTGTTGACA
AGATCCGATAAAAATAGAGGGAAGAGTGATAACGTCCCCTCAGAAGAAGTTGT
CAAGAAAATGAAAAATTATTGGCGGCAGCTGCTGAACGCCAAACTGATCACAC
AACGGAAGTTCGATAATCTGACTAAGGCTGAACGAGGTGGCCTGTCTGAGTTG
GATAAAGCCGGCTTCATCAAAAGGCAGCTTGTTGAGACACGCCAGATCACCAA
GCACGTGGCCCAAATTCTCGATTCACGCATGAACACCAAGTACGATGAAAATG
ACAAACTGATTCGAGAGGTGAAAGTTATTACTCTGAAGTCTAAGCTGGTCTCAG
ATTTCAGAAAGGACTTTCAGTTTTATAAGGTGAGAGAGATCAACAATTACCACC
ATGCGCATGATGCCTACCTGAATGCAGTGGTAGGCACTGCACTTATCAAAAAAT
ATCCCAAGCTTGAATCTGAATTTGTTTACGGAGACTATAAAGTGTACGATGTTA
GGAAAATGATCGCAAAGTCTGAGCAGGAAATAGGCAAGGCCACCGCTAAGTAC
TTCTTTTACAGCAATATTATGAATTTTTTCAAGACCGAGATTACACTGGCCAAT
GGAGAGATTCGGAAGCGACCACTTATCGAAACAAACGGAGAAACAGGAGAAA
TCGTGTGGGACAAGGGTAGGGATTTCGCGACAGTCCGGAAGGTCCTGTCCATG
CCGCAGGTGAACATCGTTAAAAAGACCGAAGTACAGACCGGAGGCTTCTCCAA
GGAAAGTATCCTCCCGAAAAGGAACAGCGACAAGCTGATCGCACGCAAAAAA
GATTGGGACCCCAAGAAATACGGCGGATTCGATTCTCCTACAGTCGCTTACAGT
GTACTGGTTGTGGCCAAAGTGGAGAAAGGGAAGTCTAAAAAACTCAAAAGCGT
CAAGGAACTGCTGGGCATCACAATCATGGAGCGATCAAGCTTCGAAAAAAACC
CCATCGACTTTCTCGAGG
CGAAAGGATATAAAGAGGTCAAAAAAGACCTCATCATTAAGCTTCCCAAGTAC
TCTCTCTTTGAGCTTGAAAACGGCCGGAAACGAATGCTCGCTAGTGCGGGCGA
GCTGCAGAAAGGTAACGAGCTGGCACTGCCCTCTAAATACGTTAATTTCTTGTA
TCTGGCCAGCCACTATGAAAAGCTCAAAGGGTCTCCCGAAGATAATGAGCAGA
AGCAGCTGTTCGTGGAACAACACAAACACTACCTTGATGAGATCATCGAGCAA
ATAAGCGAATTCTCCAAAAGAGTGATCCTCGCCGACGCTAACCTCGATAAGGT

GCTTTCTGCTTACAATAAGCACAGGGATAAGCCCATCAGGGAGCAGGCAGAAA
ACATTATCCACTTGTTTACTCTGACCAACTTGGGCGCGCCTGCAGCCTTCAAGT
ACTTCGACACCACCATAGACAGAA
AGCGGTACACCTCTACAAAGGAGGTCCTGGACGCCACACTGATTCATCAGTCA
ATTACGGGGCTCTATGAAACAAGAATCGACCTCTCTCAGCTCGGTGGAGACAG
CAGGGCTGACCCCAAGAAGAAGAGGAAGGTGAGGGGAGGCGGAGGATCTGGC
GGAGGCGGAAGTGGCGGAGGGGGCAGCATGAACCCTAGCATGAAGCAGAAGC
AGGAAGAGATCAAAGAGAACATCAAGAACAGCAGCGTGCCCAGACGGACCCT
GAAGATGATCCAGCCTAGCGCCAGCGGCAGCCTCGTGGGCAGAGAGAATGAAC
TGTCTGCCGGCCTGAGCAAGCGGAAGCACAGAAACGACCACCTGACCAGCACC
ACCAGCAGCCCTGGCGTGATCGTGCCTGAGAGCAGCGAGAACAAGAACCTGGG
CGGCGTGACCCAGGAATCCTTCGACCTGATGATCAAAGAAAACCCCAGCAGCC
AGTATTGGAAAGAGGTGGCCGAGAAGCGGCGGAAGGCCCTG (SEQ ID NO: 25)

Cdt1:linker:Cas9

ATGCCATCTCCTGCCAGACCTGCTCTGAGAGCCCCTGCCTCTGCCACAAGCGGC
AGCAGAAAGAGAGCCAGACCTCCTGCCGCCCCTGGCAGAGATCAGGCTAGACC
TCCAGCTCGGCGGAGACTGAGACTGAGCGTGGACGAGGTGTCCAGCCCTAGCA
CACCTGAGGCCCCTGATATCCCCGCCTGTCCTAGCCCTGGCCAGAAGATCAAGA
AGTCCACCCCTGCCGCCGGACAGCCTCCTCATCTGACATCTGCCCAGGACCAGG
ACACCATCGGAGGCGGAGGATCTGGCGGAGGCGGAAGTGGCGGAGGGGGCAG
CGACAAGAAGTACTCCATTGGGCTCGATATCGGCACAAACAGCGTCGGCTGGG
CCGTCATTACGGACGAGTACAAGGTGCCGAGCAAAAAATTCAAAGTTCTGGGC
AATACCGATCGCCACAGCATAAA
GAAGAACCTCATTGGCGCCCTCCTGTTCGACTCCGGGGAGACGGCCGAAGCCA
CGCGGCTCAAAAGAACAGCACGGCGCAGATATACCCGCAGAAAGAATCGGATC
TGCTACCTGCAGGAGATCTTTAGTAATGAGATGGCTAAGGTGGATGACTCTTTC
TTCCATAGGCTGGAGGAGTCCTTTTTGGTGGAGGAGGATAAAAAGCACGAGCG
CCACCCAATCTTTGGCAATATCGTGGACGAGGTGGCGTACCATGAAAAGTACC
CAACCATATATCATCTGAGGAAGAAGCTTGTAGACAGTACTGATAAGGCTGAC
TTGCGGTTGATCTATCTCGCGCTGGCGCATATGATCAAATTTCGGGGACACTTC
CTCATCGAGGGGGACCTGAACCCAGACAACAGCGATGTCGACAAACTCTTTAT
CCAACTGGTTCAGACTTACAATCAGCTTTTCGAAGAGAACCCGATCAACGCATC
CGGAGTTGACGCCAAAGCAATCCTGAGCGCTAGGCTGTCCAAATCCCGGCGGC

TCGAAAACCTCATCGCACAGCTCCCTGGGGAGAAGAAGAACGGCCTGTTTGGT
AATCTTATCGCCCTGTCACTCGGGCTGACCCCCAACTTTAAATCTAACTTCGAC
CTGGCCGAAGATGCCAAGCTTCAACTGAGCAAAGACACCTACGATGATGATCT
CGACAATCTGCTGGCCCAGATCGGCGACCAGTACGCAGACCTTTTTTTGGCGGC
AAAGAACCTGTCAGACGCCATTCTGCTGAGTGATATTCTGCGAGTGAACACGG
AGATCACCAAAGCTCCGCTGAGCGCTAGTATGATCAAGCGCTATGATGAGCAC
CACCAAGACTTGACTTTGCTGAAGGCCCTTGTCAGACAGCAACTGCCTGAGAA
GTACAAGGAAATTTTCTTCGATCAGTCTAAAAATGGCTACGCCGGATACATTGA
CGGCGGAGCAAGCCAGGAGGAATTTTACAAATTTATTAAGCCCATCTTGGAAA
AAATGGACGGCACCGAGGAGCTGCTGGTAAAGCTTAACAGAGAAGATCTGTTG
CGCAAACAGCGCACTTTCGACAATGGAAGCATCCCCCACCAGATTCACCTGGG
CGAACTGCACGCTATCCTCAGGCGGCAAGAGGATTTCTACCCCTTTTTGAAAGA
TAACAGGGAAAAGATTGAGAAAATCCTCACATTTCGGATACCCTACTATGTAG
GCCCCCTCGCCCGGGGAAATTCCAGATTCGCGTGGATGACTCGCAAATCAGAA
GAGACCATCACTCCCTGGAACTTCGAGGAAGTCGTGGATAAGGGGGCCTCTGC
CCAGTCCTTCATCGAAAGGATGACTAACTTTGATAAAAATCTGCCTAACGAAAA
GGTGCTTCCTAAACACTCTCTGCTGTACGAGTACTTCACAGTTTATAACGAGCT
CACCAAGGTCAAATACGTCACAGAAGGGATGAGAAAGCCAGCATTCCTGTCTG
GAGAGCAGAAGAAAGCTATCGTGGACCTCCTCTTCAAGACGAACCGGAAAGTT
ACCGTGAAACAGCTCAAAGAAGACTATTTCAAAAAGATTGAATGTTTCGACTCT
GTTGAAATCAGCGGAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGTATCA
CGATCTCCTGAAAATCATTAAAGACAAGGACTTCCTGGACAATGAGGAGAACG
AGGACATTCTTGAGGACATTGTCCTCACCCTTACGTTGTTTGAAGATAGGGAGA
TGATTGAAGAACGCTTGAAAACTTACGCTCATCTCTTCGACGACAAAGTCATGA
AACAGCTCAAGAGGCGCCGATATACAGGATGGGGGCGGCTGTCAAGAAAACTG
ATCAATGGGATCCGAGACAAGCAGAGTGGAAAGACAATCCTGGATTTTCTTAA
GTCCGATGGATTTGCCAACCGGAACTTCATGCAGTTGATCCATGATGACTCTCT
CACCTTTAAGGAGGACATCCAGAAAGCACAAGTTTCTGGCCAGGGGGACAGTC
TTCACGAGCACATCGCTAATCTTGCAGGTAGCCCAGCTATCAAAAAGGGAATA
CTGCAGACCGTTAAGGTCGTGGATGAACTCGTCAAAGTAATGGGAAGGCATAA
GCCCGAGAATATCGTTATCGAGATGGCCCGAGAGAACCAAACTACCCAGAAGG
GACAGAAGAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGTATAAAAGA
ACTGGGGTCCCAAATCCTTAAGGAACACCCAGTTGAAAACACCCAGCTTCAGA

ATGAGAAGCTCTACCTGTACTACCTGCAGAACGGCAGGGACATGTACGTGGAT
CAGGAACTGGACATCAATCGGCTCTCCGACTACGACGTGGATCATATCGTGCCC
CAGTCTTTTCTCAAAGATGATTCTATTGATAATAAAGTGTTGACAAGATCCGAT
AAAAATAGAGGGAAGAGTGATAACGTCCCCTCAGAAGAAGTTGTCAAGAAAAT
GAAAAATTATTGGCGGCAGCTGCTGAACGCCAAACTGATCACACAACGGAAGT
TCGATAATCTGACTAAGGCTGAACGAGGTGGCCTGTCTGAGTTGGATAAAGCC
GGCTTCATCAAAAGGCAGCTTGTTGAGACACGCCAGATCACCAAGCACGTGGC
CCAAATTCTCGATTCACGCATGAACACCAAGTACGATGAAAATGACAAACTGA
TTCGAGAGGTGAAAGTTATTACTCTGAAGTCTAAGCTGGTCTCAGATTTCAGAA
AGGACTTTCAGTTTTATAAGGTGAGAGAGATCAACAATTACCACCATGCGCATG
ATGCCTACCTGAATGCAGTGGTAGGCACTGCACTTATCAAAAAATATCCCAAGC
TTGAATCTGAATTTGTTTACGGAGACTATAAAGTGTACGATGTTAGGAAAATGA
TCGCAAAGTCTGAGCAGGAAATAGGCAAGGCCACCGCTAAGTACTTCTTTTAC
AGCAATATTATGAATTTTTTCAAGACCGAGATTACACTGGCCAATGGAGAGATT
CGGAAGCGACCACTTATCGAAACAAACGGAGAAACAGGAGAAATCGTGTGGG
ACAAGGGTAG

GGATTTCGCGACAGTCCGGAAGGTCCTGTCCATGCCGCAGGTGAACATCGTTAA
AAAGACCGAAGTACAGACCGGAGGCTTCTCCAAGGAAAGTATCCTCCCGAAAA
GGAACAGCGACAAGCTGATCGCACGCAAAAAAGATTGGGACCCCAAGAAATA
CGGCGGATTCGATTCTCCTACAGTCGCTTACAGTGTACTGGTTGTGGCCAAAGT
GGAGAAAGGGAAGTCTAAAAAACTCAAAAGCGTCAAGGAACTGCTGGGCATC
ACAATCATGGAGCGATCAAGCTTCGAAAAAAACCCCATCGACTTTCTCGAGGC
GAAAGGATATAAAGAGGTCAAAAAAGACCTCATCATTAAGCTTCCCAAGTACT
CTCTCTTTGAGCTTGAAAACGGCCGGAAACGAATGCTCGCTAGTGCGGGCGAG
CTGCAGAAAGGTAACGAGCTGGCACTGCCCTCTAAATACGTTAATTTCTTGTAT
CTGGCCAGCCACTATGAAAAGCTCAAAGGGTCTCCCGAAGATAATGAGCAGAA
GCAGCTGTTCGTGGAACAACACAAACACTACCTTGATGAGATCATCGAGCAAA
TAAGCGAATTCTCCAAAAGAGTGATCCTCGCCGACGCTAACCTCGATAAGGTG
CTTTCTGCTTACAATAAGCACAGGGATAAGCCCATCAGGGAGCAGGCAGAAAA
CATTATCCACTTGTTTACTCTGACCAACTTGGGCGCGCCTGCAGCCTTCAAGTA
CTTCGACACCACCATAGACAGAAAGCGGTACACCTCTACAAAGGAGGTCCTGG
ACGCCACACTGATTCATCAGTCAATTACGGGGCTCTATGAAACAAGAATCGAC

CTCTCTCAGCTCGGTGGAGACAGCAGGGCTGACCCCAAGAAGAAGAGGAAGGT
GAGG (SEQ ID NO: 26)

Gem:linker:Cas9

ATGAACCCTAGCATGAAGCAGAAGCAGGAAGAGATCAAAGAGAACATCAAGA
ACAGCAGCGTGCCCAGACGGACCCTGAAGATGATCCAGCCTAGCGCCAGCGGC
AGCCTCGTGGGCAGAGAGAATGAACTGTCTGCCGGCCTGAGCAAGCGGAAGCA
CAGAAACGACCACCTGACCAGCACCACCAGCAGCCCTGGCGTGATCGTGCCTG
AGAGCAGCGAGAACAAGAACCTGGGCGGCGTGACCCAGGAATCCTTCGACCTG
ATGATCAAAGAAAACCCCAGCAGCCAGTATTGGAAAGAGGTGGCCGAGAAGC
GGCGGAAGGCCCTGGGAGGCGGAGGATCTGGCGGAGGCGGAAGTGGCGGAGG
GGGCAGCGACAAGAAGTACTCCATTGGGCTCGATATCGGCACAAACAGCGTCG
GCTGGGCCGTCATTACGGACGAG
TACAAGGTGCCGAGCAAAAAATTCAAAGTTCTGGGCAATACCGATCGCCACAG
CATAAAGAAGAACCTCATTGGCGCCCTCCTGTTCGACTCCGGGGAGACGGCCG
AAGCCACGCGGCTCAAAAGAACAGCACGGCGCAGATATACCCGCAGAAAGAA
TCGGATCTGCTACCTGCAGGAGATCTTTAGTAATGAGATGGCTAAGGTGGATGA
CTCTTTCTTCCATAGGCTGGAGGAGTCCTTTTTGGTGGAGGAGGATAAAAAGCA
CGAGCGCCACCCAATCTTTGGCAATATCGTGGACGAGGTGGCGTACCATGAAA
AGTACCCAACCATATATCATCTGAGGAAGAAGCTTGTAGACAGTACTGATAAG
GCTGACTTGCGGTTGATCTATCTCGCGCTGGCGCATATGATCAAATTTCGGGGA
CACTTCCTCATCGAGGGGGACCTGAACCCAGACAACAGCGATGTCGACAAACT
CTTTATCCAACTGGTTCAGACTTACAATCAGCTTTTCGAAGAGAACCCGATCAA
CGCATCCGGAGTTGACGCCAAAGCAATCCTGAGCGCTAGGCTGTCCAAATCCC
GGCGGCTCGAAAACCTCATCGCACAGCTCCCTGGGGAGAAGAAGAACGGCCTG
TTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACCCCCAACTTTAAATCTAACT
TCGACCTGGCCGAAGATGCCAAGCTTCAACTGAGCAAAGACACCTACGATGAT
GATCTCGACAATCTGCTGGCCCAGATCGGCGACCAGTACGCAGACCTTTTTTTG
GCGGCAAAGAACCTGTCAGACGCCATTCTGCTGAGTGATATTCTGCGAGTGAA
CACGGAGATCACCAAAGCTCCGCTGAGCGCTAGTATGATCAAGCGCTATGATG
AGCACCACCAAGACTTGACTTTGCTGAAGGCCCTTGTCAGACAGCAACTGCCTG
AGAAGTACAAGGAAATTTTCTTCGATCAGTCTAAAAATGGCTACGCCGGATAC
ATTGACGGCGGAGCAAGCCAGGAGGAATTTTACAAATTTATTAAGCCCATCTTG
GAAAAAATGGACGGCACCGAGGAGCTGCTGGTAAAGCTTAACAGAGAAGATCT

GTTGCGCAAACAGCGCACTTTCGACAATGGAAGCATCCCCCACCAGATTCACCT
GGGCGAACTGCACGCTATCCTCAGGCGGCAAGAGGATTTCTACCCCTTTTTGAA
AGATAACAGGGAAAAGATTGAGAAAATCCTCACATTTCGGATACCCTACTATG
TAGGCCCCCTCGCCCGGGGAAATTCCAGATTCGCGTGGATGACTCGCAAATCA
GAAGAGACCATCACTCCCTGGAACTTCGAGGAAGTCGTGGATAAGGGGGCCTC
TGCCCAGTCCTTCATCGAAAGGATGACTAACTTTGATAAAAATCTGCCTAACGA
AAAGGTGCTTCCTAAACACTCTCTGCTGTACGAGTACTTCACAGTTTATAACGA
GCTCACCAAGGTCAAATACGTCACAGAAGGGATGAGAAAGCCAGCATTCCTGT
CTGGAGAGCAGAAGAAAGCTATCGTGGACCTCCTCTTCAAGACGAACCGGAAA
GTTACCGTGAAACAGCTCAAAGAAGACTATTTCAAAAAGATTGAATGTTTCGA
CTCTGTTGAAATCAGCGGAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGT
ATCACGATCTCCTGAAAATCATTAAAGACAAGGACTTCCTGGACAATGAGGAG
AACGAGGACATTCTTGAGGACATTGTCCTCACCCTTACGTTGTTTGAAGATAGG
GAGATGATTGAAGAACGCTTGAAAACTTACGCTCATCTCTTCGACGACAAAGTC
ATGAAACAGCTCAAGAGGCGCCGATATACAGGATGGGGGCGGCTGTCAAGAA
AACTGATCAATGGGATCCGAGACAAGCAGAGTGGAAAGACAATCCTGGATTTT
CTTAAGTCCGATGGATTTGCCAACCGGAACTTCATGCAGTTGATCCATGATGAC
TCTCTCACCTTTAAGGAGGACATCCAGAAAGCACAAGTTTCTGGCCAGGGGGA
CAGTCTTCACGAGCACATCGCTAATCTTGCAGGTAGCCCAGCTATCAAAAAGG
GAATACTGCAGACCGTTAAGGTCGTGGATGAACTCGTCAAAGTAATGGGAAGG
CATAAGCCCGAGAATATCGTTATCGAGATGGCCCGAGAGAACCAAACTACCCA
GAAGGGACAGAAGAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGTATA
AAAGAACTGGGGTCCCAAATCCTTAAGGAACACCCAGTTGAAAACACCCAGCT
TCAGAATGAGAAGCTCTACCTGTACTACCTGCAGAACGGCAGGGACATGTACG
TGGATCAGGAACTGGACATCAATCGGCTCTCCGACTACGACGTGGATCATATCG
TGCCCCAGTCTTTTCTCAAAGATGATTCTATTGATAATAAAGTGTTGACAAGAT
CCGATAAAAATAGAGGGAAGAGTGATAACGTCCCCTCAGAAGAAGTTGTCAAG
AAAATGAAAAATTATTGGCGGCAGCTGCTGAACGCCAAACTGATCACACAACG
GAAGTTCGATAATCTGACTAAGGCTGAACGAGGTGGCCTGTCTGAGTTGGATA
AAGCCGGCTTCATCAAAAGGCAGCTTGTTGAGACACGCCAGATCACCAAGCAC
GTGGCCCAAATTCTCGATTCACGCATGAACACCAAGTACGATGAAAATGACAA
ACTGATTCGAGAGGTGAAAGTTATTACTCTGAAGTCTAAGCTGGTCTCAGATTT
CAGAAAGGACTTTCAGTTTTATAAGGTGAGAGAGATCAACAATTACCACCATG

CGCATGATGCCTACCTGAATGCAGTGGTAGGCACTGCACTTATCAAAAAATATC
CCAAGCTTGAATCTGAATTTGTTTACGGAGACTATAAAGTGTACGATGTTAGGA
AAATGATCGCAAAGTCTGAGCAGGAAATAGGCAAGGCCACCGCTAAGTACTTC
TTTTACAGCAATATTATGAATTTTTTCAAGACCGAGATTACACTGGCCAATGGA
GAGATTCGGA
AGCGACCACTTATCGAAACAAACGGAGAAACAGGAGAAATCGTGTGGGACAA
GGGTAGGGATTTCGCGACAGTCCGGAAGGTCCTGTCCATGCCGCAGGTGAACA
TCGTTAAAAAGACCGAAGTACAGACCGGAGGCTTCTCCAAGGAAAGTATCCTC
CCGAAAAGGAACAGCGACAAGCTGATCGCACGCAAAAAAGATTGGGACCCCA
AGAAATACGGCGGATTCGATTCTCCTACAGTCGCTTACAGTGTACTGGTTGTGG
CCAAAGTGGAGAAAGGGAAGTCTAAAAAACTCAAAAGCGTCAAGGAACTGCT
GGGCATCACAATCATGGAGCGATCAAGCTTCGAAAAAAACCCCATCGACTTTC
TCGAGGCGAAAGGATATAAAGAGGTCAAAAAAGACCTCATCATTAAGCTTCCC
AAGTACTCTCTCTTTGAGCTTGAAAACGGCCGGAAACGAATGCTCGCTAGTGCG
GGCGAGCTGCAGAAAGGTAACGAGCTGGCACTGCCCTCTAAATACGTTAATTT
CTTGTATCTGGCCAGCCACTATGAAAAGCTCAAAGGGTCTCCCGAAGATAATG
AGCAGAAGCAGCTGTTCGTGGAACAACACAAACACTACCTTGATGAGATCATC
GAGCAAATAAGCGAATTCTCCAAAAGAGTGATCCTCGCCGACGCTAACCTCGA
TAAGGTGCTTTCTGCTTACAATAAGCACAGGGATAAGCCCATCAGGGAGCAGG
CAGAAAACATTATCCACTTGTTTACTCTGACCAACTTGGGCGCGCCTGCAGCCT
TCAAGTACTTCGACACCACCATAGACAGAAAGCGGTACACCTCTACAAAGGAG
GTCCTGGACGCCACACTGATTCATCAGTCAATTACGGGGCTCTATGAAACAAGA
ATCGACCTCTCTCAGCTCGGTGGAGACAGCAGGGCTGACCCCAAGAAGAAGAG
GAAGGTGAGG (SEQ ID NO: 27)

EXAMPLE 4

Auto-regulation of RNA-guided nucleases by engineering partial target sites

Reducing RNA-directed nuclease activity can be beneficial towards reducing the potential for off-target effects. In a clinical setting, this would be highly significant with viral delivery methods, such as AAV, which are characterized by prolonged or life-time expression.

Here, a self-regulating system is described for the expression of Cas9, an RNA-guided nuclease, that relies upon the use of imperfect target sites encoded within the Cas9/gRNA expression system itself. Cas9 binding (without cleavage) has been shown by

numerous studies to be effective at regulating gene expression. Most studies use the nuclease-dead version of Cas9 to prevent cleavage activity, however a system which allows for both cleavage and self-regulation would be highly desirable.

By using engineered sequences that correspond to partial target sites, the cutting and binding activities of Cas9 can be separated; partial sequence complementarity allows Cas9 to bind without DNA cleavage. Using this binding propensity, Cas9 can be directed back to regulating its own expression. For SpCas9, sites that are generally below 17nt of complementarity do not result in cleavage, and even single mismatches can result in no cleavage with high-fidelity or high-specificity Cas9 mutants.

Importantly, all of this regulation occurs with minimal sequence changes, and without additional binding factors, fusions, or other proteins, keeping the expression cassette size small enough to be delivered by AAV.

Methods

Plasmid construction: The GFP reporter plasmids were constructed by Gibson Assembly (New England Biolabs, Ipswich, MA) (Gibson *et al.* (2009) *Nature Methods* 6:343-345) with slight modifications.

Human embryonic kidney (HEK) cell line 293T (Life Technologies, Grand Island, NY) was maintained at 37°C with 5% CO_2 / 20% O_2 in Dulbecco's modified Eagle's Medium (DMEM) (Invitrogen) supplemented with 10% fetal bovine serum (heat-inactivated) (Gibco, Life Technologies, Grand Island, NY) and 2mM GlutaMAX (Invitrogen).

N2A cell line (Life Technologies, Grand Island, NY) was maintained at 37°C with 5% CO_2 / 20% O_2 in Dulbecco's modified Eagle's Medium (DMEM) (Invitrogen) supplemented with 10% fetal bovine serum (Gibco, Life Technologies, Grand Island, NY) and 2mM GlutaMAX (Invitrogen).

Cells were transfected (Lipofectamine 3000) with each construct. GFP expression was analyzed using flow cytometry.

mm079 target sequence: GAAGAAGGTTCGAGATCTCA (SEQ ID NO: 28)

mm079 genomic target site: GAAGAAGGTTCGAGATCTCAAGG (SEQ ID NO: 29)

Auto1 site: GTTCGAGATCTCAGGGAAT (SEQ ID NO: 30)

Auto2 site: GTTCGAGATCTCAGGGTTT (SEQ ID NO: 31)

H1_wt:

GGAATTCGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAGTG

TCACTAGGCGGGAACACCCAGCGCGCGTGCGCCCTGGCAGGAAGATGGCTGTG
AGGGACAGGGGAGTGGCGCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGA
AATCACCATAAACGTGAAATGTCTTTGGATTTGGGAATCTTATAAGTTCTGTAT
GAGACCA CTTTTTCCC (SEQ ID NO: 32)

H1_Auto1:

GGAATTCGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAGTG
TCACTAGGCGGGAACACCCAGCGCGCGTGCGCCCTGGCAGGAAGATGGCTGTG
AGGGACAGGGGAGTGGCGCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGA
AATCACCATAAACGTGAAAGTTCGAGATCTCAGGGAATCTTATAAGTTCTGTAT
GAGACC ACTTTTTCCC (SEQ ID NO: 33)

H1_Auto2:

GGAATTCGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAGTG
TCACTAGGCGGGAACACCCAGCGCGCGTGCGCCCTGGCAGGAAGATGGCTGTG
AGGGACAGGGGAGTGGCGCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGA
AATCACCATAAACGTGAAATGTCTTTGGATTTGGGAATCTTATAAGTGTTCGAG
ATCTCAG GGTTTTCCC (SEQ ID NO: 34)

EXAMPLE 5

Methionine (Met) followed by a glycine (Gly), proline (Pro), serine (Ser), threonine (Thr), alanine (Ala), valine (Val), or cysteine (Cys) are processed by Methionineaminopeptidases, which cleave off the N-terminal met (FIG. 1A). The identity of the position 2 amino-acid then either stabilizes or destabilizes the entire protein. Thus, by changing the identity of the second amino acid, one can alter some proteins half-life through Met-aminopeptidases.

While differing from organism to organism, and variable between proteins, the N-end rule serves as an estimation of a proteins half-life within a cell (FIG. 17B). Protein levels are modulated through the ubiquitin (Ub) proteolytic system. Ubiquitin is a highly conserved protein of 76 amino acids that is typically associated with protein degradation; Ub molecules are conjugated to target proteins marking those proteins for destruction by the proteasome.

Linear poly-ubiquitin is recognized in the cell by deubiquitin enzymes which cleave the individual Ub peptides. This process can be co-opted to generate specific N-terminal residues by fusing Ub to the N-terminus of any protein (FIG. 18). Once inside the cell, deubiquitin enzymes recognize and cleave the Ub moiety, releasing the fused protein. This

process can be used to generate precise N-terminal amino-acid residues on a given protein, the identity of which determines the proteins half-life.

Generically, controlling the expression of either the RNA-guided nuclease or the guide RNA can modulate the entire holoenzyme complex; for the CRISPR-Cas9 system, this can be done either by regulating the levels of Cas9 or the gRNA. Furthermore, simple changes in the N-terminal amino acids, either through Met-aminopeptidases, deubiquintation, or alternative methods, can accomplish this regulation (FIG. 19A).

Reducing Cas9 half-life can be beneficial towards reducing the potential for off-target effects. In a clinical setting, this would be highly significant with viral delivery methods, such as AAV, which are characterized by prolonged or life-time expression. Conversely, with inefficient methods of delivery, increased protein half-life could be desirable or necessary to elicit an effect. In various diverse settings, the ability to tune the levels of the nuclease would be highly desirable. Additionally, this approach to regulation is simple and can be used in conjunction with other forms of regulation, such as degrons, inducible degrons, or split variants of RNA-guided nucleases.

Here, it was shown that levels of Cas9, an RNA-guided nuclease, can be modulated by the identity of the N-terminal amino acid, by as much as 8-10-fold, a significant range of expression for a nuclease (FIG. 19A). Also shown is how commonly used 2A peptides can be leveraged to modulate protein levels in the cell (FIG. 20).

Methods

Plasmid construction: To generate the ubiquitin fused Cas9 proteins, ubiquitin was fused to the N-terminus of Cas9 using Gibson Assembly (New England Biolabs, Ipswich, MA) (Gibson *et al.* (2009) *Nature Methods* 6:343-345) with slight modifications. Briefly, primers encoding all 20 amino acids were used to amplify a gBlock encoding ubiquitin, which was then cloned in-frame with Cas9. All constructs were sequenced verified.

Human embryonic kidney (HEK) cell line 293T (Life Technologies, Grand Island, NY) was maintained at 37°C with 5% CO_2 / 20% O_2 in Dulbecco's modified Eagle's Medium (DMEM) (Invitrogen) supplemented with 10% fetal bovine serum (Gibco, Life Technologies, Grand Island, NY) and 2mM GlutaMAX (Invitrogen).

Cells were transfected with each construct at either 2000ng, 500ng, or 125ng. 48hrs later cells were harvested for protein and then analyzed on a Simple Simon machine for

Cas9 expression or GAPDH for control. The ratios were used to normalize values and to determine protein stability.

1. Tasaki T *et al.* (2012) *Annu Rev Biochem*. 81:261-89.

2. Varshavsky A *et al.* (2011) *Protein Sci*. 20(8):1298-345.

3. Bachmair A *et al.* (1986) *Science*. 234(4773):179-86.

Table 1: Approximate half-life of proteins in mammalian cells based on the N-terminal residue

Val	100hr
Met	30hr
Gly	30hr
Pro	20hr
Ile	20hr
Thr	7.2hr
Leu	5.5hr
Ala	4.4hr
His	3.5hr
Trp	2.8hr
Tyr	2.8hr
Ser	1.9hr
Asn	1.4hr
Lys	1.3hr
Cys	1.2hr
Asp	1.1hr
Phe	1.1hr
Glu	1.0hr
Arg	1.0hr
Gln	0.8hr

Table 2: List of commonly used 2A sequences

P2A: GSGATNFSLLKQAGDVEENPGP (SEQ ID NO: 35)

T2A: GSGEGRGSLLTCGDVEENPGP (SEQ ID NO: 36)

E2A: GSGQCTNYALLKLAGDVESNPGP (SEQ ID NO: 37)

F2A: GSGVKQTLNFDLLKLAGDVESNPGP (SEQ ID NO: 38)

Protein Sequences

Ub:Val::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSD
YNIQKESTLHLVLRLRGGVDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD
DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL
RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA
KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD

NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKG
QKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ
ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII
EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 39)

Ub:Met::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSD
YNIQKESTLHLVLRLRGG**M**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD
DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL
RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA
KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD
NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG

DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKG
QKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYW
RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRM
NTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG
TALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSK
ESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKE
LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGEL
QKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
RYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ ID NO: 40)

Ub:Gly::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSD
YNIQKESTLHLVLRLRGG**G**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD
DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL
RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA
KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD
NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKG
QKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ
ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN

FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII
EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV(SEQ ID NO: 41)

Ub:Pro::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSD
YNIQKESTLHLVLRLRGG**P**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD
DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL
RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA
KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD
NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKG
QKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ
ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII
EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF

DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ ID NO: 42)

Ub:Ile::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSD YNIQKESTLHLVLRLRGGIDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ ID NO: 43)

Ub:Thr::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSD YNIQKESTLHLVLRLRGGTDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT

DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD
DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL
RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA
KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD
NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKG
QKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ
ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII
EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 44)

Ub:Leu::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSD
YNIQKESTLHLVLRLRGGLDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD
DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL
RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA
KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI

KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD
NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKG
QKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ
ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII
EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 45)

Ub:Ala::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSD
YNIQKESTLHLVLRLRGGADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD
DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL
RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA
KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD
NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK

DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKG
QKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ
ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII
EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 46)

Ub:His::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSD
YNIQKESTLHLVLRLRGGHDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD
DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL
RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA
KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD
NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKG
QKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK

MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ
ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII
EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 47)

Ub:Trp::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSD
YNIQKESTLHLVLRLRGGWDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGN
TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD
DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL
RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA
KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD
NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKG
QKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ
ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK

LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII
EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 48)

Ub:Try::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSD
YNIQKESTLHLVLRLRGGYDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD
DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL
RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA
KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD
NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKG
QKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ
ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII
EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 49)

Ub:Ser::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSD
YNIQKESTLHLVLRLRGGSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD
DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL
RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA
KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD
NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKG
QKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ
ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII
EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 50)

Ub:Asn::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSD
YNIQKESTLHLVLRLRGGNDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD
DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL

RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA
KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD
NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKG
QKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ
ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII
EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 51)

Ub:Lys::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSD
YNIQKESTLHLVLRLRGG**K**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD
DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL
RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA
KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD

NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKG
QKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ
ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII
EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 52)

Ub:Cys::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSD
YNIQKESTLHLVLRLRGG**C**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD
DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL
RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA
KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD
NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG

DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKG
QKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ
ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII
EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 53)

Ub:Asp::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSD
YNIQKESTLHLVLRLRGG**D**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD
DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL
RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA
KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD
NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKG
QKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ
ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL

NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII
EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 54)

Ub:Phe::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSD
YNIQKESTLHLVLRLRGGFDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD
DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL
RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA
KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD
NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKG
QKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ
ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII

EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 55)

Ub:Glu::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSD
YNIQKESTLHLVLRLRGGEDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD
DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL
RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA
KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD
NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKG
QKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ
ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII
EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 56)

Ub:Arg::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSD
YNIQKESTLHLVLRLRGG**R**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD
DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL
RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA
KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD
NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKG
QKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ
ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII
EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 57)

Ub:Gln::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSD
YNIQKESTLHLVLRLRGG**Q**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD
DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL
RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA

KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD
NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKG
QKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ
ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII
EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 58)

Ub(G76V)::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSD
YNIQKESTLHLVLRLRG**V**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD
RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDS
FFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRL
IYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKA
ILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSK
DTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK
MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNRE
KIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM

TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLS
RKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK
NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDIN
RLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ
ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII
EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 59)

DNA Sequences

Ub:

ACCACCTCTCAGACGCAGGACCAGGTGCAGGGTCGACTCCTTCTGGATGTTGTA
GTCAGAAAGAGTGCGGCCATCTTCCAGCTGCTTGCCTGCAAAGATGAGCCTCTG
CTGGTCGGGAGGGATGCCTTCTTTATCCTGGATCTTGGCCTTCACATTTTCGATG
GTGTCACTGGGCTCCACTTCCAGGGTGATGGTCTTGCCGGTCAGGGTCTTCACG
AAGATCTGCAT (SEQ ID NO: 60)

Cas9:

GACAAGAAGTACTCCATTGGGCTCGATATCGGCACAAACAGCGTCGGCTGGGC
CGTCATTACGGACGAGTACAAGGTGCCGAGCAAAAAATTCAAAGTTCTGGGCA
ATACCGATCGCCACAGCATAAAGAAGAACCTCATTGGCGCCCTCCTGTTCGACT
CCGGGGAGACGGCCGAAGCCACGCGGCTCAAAAGAACAGCACGGCGCAGATA
TACCCGCAGAAAGAATCGGATCTGCTACCTGCAGGAGATCTTTAGTAATGAGA
TGGCTAAGGTGGATGACTCTTTCTTCCATAGGCTGGAGGAGTCCTTTTTGGTGG
AGGAGGATAAAAAGCACGAGCGCCACCCAATCTTTGGCAATATCGTGGACGAG
GTGGCGTACCATGAAAAGTACCCAACCATATATCATCTGAGGAAGAAGCTTGT

AGACAGTACTGATAAGGCTGACTTGCGGTTGATCTATCTCGCGCTGGCGCATAT
GATCAAATTTCGGGGACACTTCCTCATCGAGGGGGACCTGAACCCAGACAACA
GCGATGTCGACAAACTCTTTATCCAACTGGTTCAGACTTACAATCAGCTTTTCG
AAGAGAACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAATCCTGAGCGCT
AGGCTGTCCAAATCCCGGCGGCTCGAAAACCTCATCGCACAGCTCCCTGGGGA
GAAGAAGAACGGCCTGTTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACCCC
CAACTTTAAATCTAACTTCGACCTGGCCGAAGATGCCAAGCTTCAACTGAGCAA
AGACACCTACGATGATGATCTCGACAATCTGCTGGCCCAGATCGGCGACCAGT
ACGCAGACCTTTTTTTGGCGGCAAAGAACCTGTCAGACGCCATTCTGCTGAGTG
ATATTCTGCGAGTGAACACGGAGATCACCAAAGCTCCGCTGAGCGCTAGTATG
ATCAAGCGCTATGATGAGCACCACCAAGACTTGACTTTGCTGAAGGCCCTTGTC
AGACAGCAACTGCCTGAGAAGTACAAGGAAATTTTCTTCGATCAGTCTAAAAA
TGGCTACGCCGGATACATTGACGGCGGAGCAAGCCAGGAGGAATTTTACAAAT
TTATTAAGCCCATCTTGGAAAAAATGGACGGCACCGAGGAGCTGCTGGTAAAG
CTTAACAGAGAAGATCTGTTGCGCAAACAGCGCACTTTCGACAATGGAAGCAT
CCCCCACCAGATTCACCTGGGCGAACTGCACGCTATCCTCAGGCGGCAAGAGG
ATTTCTACCCCTTTTTGAAAGATAACAGGGAAAAGATTGAGAAAATCCTCACAT
TTCGGATACCCTACTATGTAGGCCCCCTCGCCCGGGGAAATTCCAGATTCGCGT
GGATGACTCGCAAATCAGAAGAGACCATCACTCCCTGGAACTTCGAGGAAGTC
GTGGATAAGGGGGCCTCTGCCCAGTCCTTCATCGAAAGGATGACTAACTTTGAT
AAAAATCTGCCTAACGAAAAGGTGCTTCCTAAACACTCTCTGCTGTACGAGTAC
TTCACAGTTTATAACGAGCTCACCAAGGTCAAATACGTCACAGAAGGGATGAG
AAAGCCAGCATTCCTGTCTGGAGAGCAGAAGAAAGCTATCGTGGACCTCCTCTT
CAAGACGAACCGGAAAGTTACCGTGAAACAGCTCAAAGAAGACTATTTCAAAA
AGATTGAATGTTTCGACTCTGTTGAAATCAGCGGAGTGGAGGATCGCTTCAACG
CATCCCTGGGAACGTATCACGATCTCCTGAAAATCATTAAAGACAAGGACTTCC
TGGACAATGAGGAGAACGAGGACATTCTTGAGGACATTGTCCTCACCCTTACGT
TGTTTGAAGATAGGGAGATGATTGAAGAACGCTTGAAAACTTACGCTCATCTCT
TCGACGACAAAGTCATGAAACAGCTCAAGAGGCGCCGATATACAGGATGGGGG
CGGCTGTCAAGAAAACTGATCAATGGGATCCGAGACAAGCAGAGTGGAAAGA
CAATCCTGGATTTTCTTAAGTCCGATGGATTTGCCAACCGGAACTTCATGCAGT
TGATCCATGATGACTCTCTCACCTTTAAGGAGGACATCCAGAAAGCACAAGTTT
CTGGCCAGGGGGACAGTCTTCACGAGCACATCGCTAATCTTGCAGGTAGCCCA

GCTATCAAAAAGGGAATACTGCAGACCGTTAAGGTCGTGGATGAACTCGTCAA
AGTAATGGGAAGGCATAAGCCCGAGAATATCGTTATCGAGATGGCCCGAGAGA
ACCAAACTACCCAGAAGGGACAGAAGAACAGTAGGGAAAGGATGAAGAGGAT
TGAAGAGGGTATAAAAGAACTGGGGTCCCAAATCCTTAAGGAACACCCAGTTG
AAAACACCCAGCTTCAGAATGAGAAGCTCTACCTGTACTACCTGCAGAACGGC
AGGGACATGTACGTGGATCAGGAACTGGACATCAATCGGCTCTCCGACTACGA
CGTGGATCATATCGTGCCCCAGTCTTTTCTCAAAGATGATTCTATTGATAATAA
AGTGTTGACAAGATCCGATAAAAATAGAGGGAAGAGTGATAACGTCCCCTCAG
AAGAAGTTGTCAAGAAAATGAAAAATTATTGGCGGCAGCTGCTGAACGCCAAA
CTGATCACACAACGGAAGTTCGATAATCTGACTAAGGCTGAACGAGGTGGCCT
GTCTGAGTTGGATAAAGCCGGCTTCATCAAAAGGCAGCTTGTTGAGACACGCC
AGATCACCAAGCACGTGGCCCAAATTCTCGATTCACGCATGAACACCAAGTAC
GATGAAAATGACAAACTGATTCGAGAGGTGAAAGTTATTACTCTGAAGTCTAA
GCTGGTCTCAGATTTCAGAAAGGACTTTCAGTTTTATAAGGTGAGAGAGATCAA
CAATTACCACCATGCGCATGATGCCTACCTGAATGCAGTGGTAGGCACTGCACT
TATCAAAAAATATCCCAAGCTTGAATCTGAATTTGTTTACGGAGACTATAAAGT
GTACGATGTTAGGAAAATGATCGCAAAGTCTGAGCAGGAAATAGGCAAGGCCA
CCGCTAAGTACTTCTTTTACAGCAATATTATGAATTTTTTCAAGACCGAGATTAC
ACTGGCCAATGGAGAGATTCGGAAGCGACCACTTATCGAAACAAACGGAGAAA
CAGGAGAAATCGTGTGGGACAAGGGTAGGGATTTCGCGACAGTCCGGAAGGTC
CTGTCCATGCCGCAGGTGAACATCGTTAAAAAGACCGAAGTACAGACCGGAGG
CTTCTCCAAGGAAAGTATCCTCCCGAAAAGGAACAGCGACAAGCTGATCGCAC
GCAAAAAAGATTGGGACCCCAAGAAATACGGCGGATTCGATTCTCCTACAGTC
GCTTACAGTGTACTGGTTGTGGCCAAAGTGGAGAAAGGGAAGTCTAAAAAACT
CAAAAGCGTCAAGGAACTGCTGGGCATCACAATCATGGAGCGATCAAGCTTCG
AAAAAAACCCCATCGACTTTCTCGAGGCGAAAGGATATAAAGAGGTCAAAAAA
GACCTCATCATTAAGCTTCCCAAGTACTCTCTCTTTGAGCTTGAAAACGGCCGG
AAACGAATGCTCGCTAGTGCGGGCGAGCTGCAGAAAGGTAACGAGCTGGCACT
GCCCTCTAAATACGTTAATTTCTTGTATCTGGCCAGCCACTATGAAAAGCTCAA
AGGGTCTCCCGAAGATAATGAGCAGAAGCAGCTGTTCGTGGAACAACACAAAC
ACTACCTTGATGAGATCATCGAGCAAATAAGCGAATTCTCCAAAAGAGTGATC
CTCGCCGACGCTAACCTCGATAAGGTGCTTTCTGCTTACAATAAGCACAGGGAT
AAGCCCATCAGGGAGCAGGCAGAAAACATTATCCACTTGTTTACTCTGACCAA

CTTGGGCGCGCCTGCAGCCTTCAAGTACTTCGACACCACCATAGACAGAAAGC
GGTACACCTCTACAAAGGAGGTCCTGGACGCCACACTGATTCATCAGTCAATTA
CGGGGCTCTATGAAACAAGAATCGACCTCTCTCAGCTCGGTGGAGACAGCAGG
GCTGACCCCAAGAAGAAGAGGAAGGTG (SEQ ID NO: 61)

EXAMPLE 6

Use of a bidirectional promoter to deliver an RNA-directed nuclease, guide RNA, and HDR templates through a single AAV virus

The delivery of RNA-directed nuclease, guide RNA, and HDR templates through a single AAV virus A revolutionary genome-editing technology known as CRISPR is transforming biological research and ushering in a new era for genetic medicine. Each cell in our body contains 3 billion base-pairs of DNA and even single changes – or mutations – can cause a wide variety of inherited or acquired diseases. Through a process akin to genome surgery, CRISPR technology allows researchers to target disease-associated mutations with unprecedented precision, treating diseases at their root cause.

Research labs across the world have shown that CRISPR can efficiently target disease mutations; however, until now the development of CRISPR-based therapeutics for human use has been hampered by delivery constraints. The preferred delivery choice for tissue-specific in vivo gene therapy is recombinant adeno-associated viruses (AAV). AAV has a notable history of safety, efficacy, and a lack of toxicity that is further illustrated by the increasing number of FDA approved trials using this approach – there have been 150 trials using AAV worldwide. These compact viruses can deliver materials with high efficiency and specificity to a wide variety of cell types, and while safe, their small size presents a significant obstacle for the packaging of CRISPR components.

Dominant diseases occur when people inherit one bad gene which then poisons the normal copy. While untreatable by gene-therapy technology, these mutations can be suppressed using CRISPR which disrupts the mutation and allows the other normal gene to work. We previously disclosed a method for the packaging of both CRISPR components into a single AAV virus using the compact H1 bidirectional promoter.

Recessive mutations, on the other hand, are the result of inheriting two bad copies of a gene. These diseases require delivering CRISPR as well as a long stretch of DNA surrounding the mutation (a template) so that the cell can edit out the mutations by copying in the correct sequence. While the vast majority of diseases fall into this category, no

means exists to deliver CRISPR components and the template DNA via a single AAV virus exist due to the size limitations of AAV.

Historically, AAV served as a method to deliver templates for site-specific DNA change. Compared to other known methods, AAV templates are the most recombinogenic, although the rates were still low and on the order of less than 1 in 104 prior to gene-editing methods. We know that DNA breaks are highly recombinogenic, and that co-delivery of CRISPR-Cas9 with an AAV template has the ability to recombine at a high frequency. We also know that ssDNA templates
as small as 40-60nt are very effective HDr templates, as is dsDNA templates 200nt and bigger. Unfortunately, there are no means to deliver the CRISPR components and the HDR template in one AAV virus.

However, because the H1 bidirectional promoter system is so compact, we can deliver both the cutting (Cas9 and gRNA) and pasting (HDR template) elements through a single AAV virus. The therapeutic potential for this class of diseases is far larger both in terms of the numbers of diseases and people that can be treated.

Here, it was shown the AAV vector design for correcting mutations *in vivo*, using the H1 promoter to express an RNA-guided nuclease (Cas9), a guide RNA, and a donor template sequence. Two examples for mutation correction are illustrated for examples 1) using the SaCas9 nickase with two gRNAs and a template correction for the rd12 mutation (FIG. 25B) and 2) the SpCas9 correction of the rd10 mutation (FIG. 27).

Methods

Human embryonic kidney (HEK) cell line 293T (Life Technologies, Grand Island, NY) was maintained at 37°C with 5% CO_2 / 20% O_2 in Dulbecco's modified Eagle's Medium (DMEM) (Invitrogen) supplemented with 10% fetal bovine serum (Gibco, Life Technologies, Grand Island, NY) and 2mM GlutaMAX (Invitrogen).

Construct design. See FIG. 28. A construct was designed to using the mouse H1 promoter to express both SpCas9 and a gRNA. Two restrictions sites were incorporated before the flanking NotI sites to allow for Gibson cloning of an HDR template: KpnI for left HDR, and HpaI for right HDR. For insertion after KpnI digestion, the following sequence 5'-
GAGAGTGCACCATAGCGGCCGCGNNNGTACCCACACAAAAAACCAACACACA G-3' (SEQ ID NO: 69), where N's represent the template region is used, or after HpaI digestion, the following sequence 5'-

GGCACCGAGTCGGTGCTTTTTTGTTNNNNNNAACGCGGCCGCCTAGAGTCGAC-3' (SEQ ID NO: 70), where N's represent the template region is used.

gRNAs (see Jaskula-Ranga, V., & Zack, D. J. (2016). grID: A CRISPR-Cas9 guide RNA Database and Resource for Genome-Editing. *bioRxiv*, 097352) were generated by overlapping oligos that were annealed and amplified by PCR using two-step amplification Phusion Flash DNA polymerase (Thermo Fisher Scientific, Rockford, IL), and subsequently purified using Zymo DNA clean and concentrator columns. The purified PCR products were then resuspended in H2O and quantitated using a NanoDrop 1000 (Thermo Fisher Scientific). The gRNA-expressing constructs were generated using the Gibson Assembly (New England Biolabs, Ipswich, MA) (Gibson *et al.* (2009) *Nature Methods* 6:343-345) with slight modifications. The total reaction volume was reduced from 20µl to 2µl. Clones were verified by Sanger sequencing.

HEK293 cells were co-transfected with Cas9 (unmodified, or cell-cycle regulated versions) and the gRNA construct targeting rhodopsin. 48hrs post transfection, genomic DNA was harvested and the sequence surrounding the target cut sites were amplified according to the primers listed in the Appendix. The PCR products were then purified and quantitated before performing the T7 Endo I assay. Briefly, 200ng of PCR product was denatured and then slowly re-annealed to allow for the formation of heteroduplexes, T7 Endonuclease I was added to the PCR products and incubated at 37C for 25 minutes to cleave heteroduplexes, the reaction was quenched in loading dye, and finally, the reaction was run on a 6% TBE PAGE gel to resolve the products. The gel was stained with SYBR-Gold, visualized, and quantitated using ImageJ. NHEJ frequencies were calculated using the binomial-derived equation:

$$1-\sqrt{1-\frac{(a+b)}{(a+b+c)}}\ X\ 100$$

; where the values of "a" and "b" are equal to the integrated area of the cleaved fragments after background subtraction and "c" is equal to the integrated area of the un-cleaved PCR product after background subtraction.

Sequences

mSPA_HDR SpCas9 entry plasmid

CCTGCTGTCTCCACCGAGCTGAGAGAGGTCGATTCTTGTTTCATAGAGCCCCGT
AATTGACTGATGAATCAGTGTGGCGTCCAGGACCTCCTTTGTAGAGGTGTACCG
CTTTCTGTCTATGGTGGTGTCGAAGTACTTGAAGGCTGCAGGCGCGCCCAAGTT
GGTCAGAGTAAACAAGTGGATAATGTTTTCTGCCTGCTCCCTGATGGGCTTATC

CCTGTGCTTATTGTAAGCAGAAAGCACCTTATCGAGGTTAGCGTCGGCGAGGAT
CACTCTTTTGGAGAATTCGCTTATTTGCTCGATGATCTCATCAAGGTAGTGTTTG
TGTTGTTCCACGAACAGCTGCTTCTGCTCATTATCTTCGGGAGACCCTTTGAGCT
TTTCATAGTGGCTGGCCAGATACAAGAAATTAACGTATTTAGAGGGCAGTGCC
AGCTCGTTACCTTTCTGCAGCTCGCCCGCACTAGCGAGCATTCGTTTCCGGCCG
TTTTCAAGCTCAAAGAGAGAGTACTTGGGAAGCTTAATGATGAGGTCTTTTTTG
ACCTCTTTATATCCTTTCGCCTCGAGAAAGTCGATGGGGTTTTTTTCGAAGCTTG
ATCGCTCCATGATTGTGATGCCCAGCAGTTCCTTGACGCTTTTGAGTTTTTTAGA
CTTCCCTTTCTCCACTTTGGCCACAACCAGTACACTGTAAGCGACTGTAGGAGA
ATCGAATCCGCCGTATTTCTTGGGGTCCCAATCTTTTTTGCGTGCGATCAGCTTG
TCGCTGTTCCTTTTCGGGAGGATACTTTCCTTGGAGAAGCCTCCGGTCTGTACTT
CGGTCTTTTTAACGATGTTCACCTGCGGCATGGACAGGACCTTCCGGACTGTCG
CGAAATCCCTACCCTTGTCCCACACGATTTCTCCTGTTTCTCCGTTTGTTTCGAT
AAGTGGTCGCTTCCGAATCTCTCCATTGGCCAGTGTAATCTCGGTCTTGAAAAA
ATTCATAATATTGCTGTAAAAGAAGTACTTAGCGGTGGCCTTGCCTATTTCCTG
CTCAGACTTTGCGATCATTTTCCTAACATCGTACACTTTATAGTCTCCGTAAACA
AATTCAGATTCAAGCTTGGGATATTTTTTGATAAGTGCAGTGCCTACCACTGCA
TTCAGGTAGGCATCATGCGCATGGTGGTAATTGTTGATCTCTCTCACCTTATAA
AACTGAAAGTCCTTTCTGAAATCTGAGACCAGCTTAGACTTCAGAGTAATAACT
TTCACCTCTCGAATCAGTTTGTCATTTTCATCGTACTTGGTGTTCATGCGTGAAT
CGAGAATTTGGGCCACGTGCTTGGTGATCTGGCGTGTCTCAACAAGCTGCCTTT
TGATGAAGCCGGCTTTATCCAACTCAGACAGGCCACCTCGTTCAGCCTTAGTCA
GATTATCGAACTTCCGTTGTGTGATCAGTTTGGCGTTCAGCAGCTGCCGCCAAT
AATTTTTCATTTTCTTGACAACTTCTTCTGAGGGGACGTTATCACTCTTCCCTCT
ATTTTTATCGGATCTTGTCAACACTTTATTATCAATAGAATCATCTTTGAGAAAA
GACTGGGGCACGATATGATCCACGTCGTAGTCGGAGAGCCGATTGATGTCCAG
TTCCTGATCCACGTACATGTCCCTGCCGTTCTGCAGGTAGTACAGGTAGAGCTT
CTCATTCTGAAGCTGGGTGTTTTCAACTGGGTGTTCCTTAAGGATTTGGGACCC
CAGTTCTTTTATACCCTCTTCAATCCTCTTCATCCTTTCCCTACTGTTCTTCTGTC
CCTTCTGGGTAGTTTGGTTCTCTCGGGCCATCTCGATAACGATATTCTCGGGCTT
ATGCCTTCCCATTACTTTGACGAGTTCATCCACGACCTTAACGGTCTGCAGTATT
CCCTTTTTGATAGCTGGGCTACCTGCAAGATTAGCGATGTGCTCGTGAAGACTG
TCCCCCTGGCCAGAAACTTGTGCTTTCTGGATGTCCTCCTTAAAGGTGAGAGAG

TCATCATGGATCAACTGCATGAAGTTCCGGTTGGCAAATCCATCGGACTTAAGA
AAATCCAGGATTGTCTTTCCACTCTGCTTGTCTCGGATCCCATTGATCAGTTTTC
TTGACAGCCGCCCCCATCCTGTATATCGGCGCCTCTTGAGCTGTTTCATGACTTT
GTCGTCGAAGAGATGAGCGTAAGTTTTCAAGCGTTCTTCAATCATCTCCCTATC
TTCAAACAACGTAAGGGTGAGGACAATGTCCTCAAGAATGTCCTCGTTCTCCTC
ATTGTCCAGGAAGTCCTTGTCTTTAATGATTTTCAGGAGATCGTGATACGTTCCC
AGGGATGCGTTGAAGCGATCCTCCACTCCGCTGATTTCAACAGAGTCGAAACAT
TCAATCTTTTTGAAATAGTCTTCTTTGAGCTGTTTCACGGTAACTTTCCGGTTCG
TCTTGAAGAGGAGGTCCACGATAGCTTTCTTCTGCTCTCCAGACAGGAATGCTG
GCTTTCTCATCCCTTCTGTGACGTATTTGACCTTGGTGAGCTCGTTATAAACTGT
GAAGTACTCGTACAGCAGAGAGTGTTTAGGAAGCACCTTTTCGTTAGGCAGATT
TTTATCAAAGTTAGTCATCCTTTCGATGAAGGACTGGGCAGAGGCCCCCTTATC
CACGACTTCCTCGAAGTTCCAGGGAGTGATGGTCTCTTCTGATTTGCGAGTCAT
CCACGCGAATCTGGAATTTCCCCGGGCGAGGGGGCCTACATAGTAGGGTATCC
GAAATGTGAGGATTTTCTCAATCTTTTCCCTGTTATCTTTCAAAAAGGGGTAGA
AATCCTCTTGCCGCCTGAGGATAGCGTGCAGTTCGCCCAGGTGAATCTGGTGGG
GGATGCTTCCATTGTCGAAAGTGCGCTGTTTGCGCAACAGATCTTCTCTGTTAA
GCTTTACCAGCAGCTCCTCGGTGCCGTCCATTTTTTCCAAGATGGGCTTAATAA
ATTTGTAAAATTCCTCCTGGCTTGCTCCGCCGTCAATGTATCCGGCGTAGCCATT
TTTAGACTGATCGAAGAAAATTTCCTTGTACTTCTCAGGCAGTTGCTGTCTGAC
AAGGGCCTTCAGCAAAGTCAAGTCTTGGTGGTGCTCATCATAGCGCTTGATCAT
ACTAGCGCTCAGCGGAGCTTTGGTGATCTCCGTGTTCACTCGCAGAATATCACT
CAGCAGAATGGCGTCTGACAGGTTCTTTGCCGCCAAAAAAAGGTCTGCGTACT
GGTCGCCGATCTGGGCCAGCAGATTGTCGAGATCATCATCGTAGGTGTCTTTGC
TCAGTTGAAGCTTGGCATCTTCGGCCAGGTCGAAGTTAGATTTAAAGTTGGGGG
TCAGCCCGAGTGACAGGGCGATAAGATTACCAAACAGGCCGTTCTTCTTCTCCC
CAGGGAGCTGTGCGATGAGGTTTTCGAGCCGCCGGGATTTGGACAGCCTAGCG
CTCAGGATTGCTTTGGCGTCAACTCCGGATGCGTTGATCGGGTTCTCTTCGAAA
AGCTGATTGTAAGTCTGAACCAGTTGGATAAAGAGTTTGTCGACATCGCTGTTG
TCTGGGTTCAGGTCCCCCTCGATGAGGAAGTGTCCCCGAAATTTGATCATATGC
GCCAGCGCGAGATAGATCAACCGCAAGTCAGCCTTATCAGTACTGTCTACAAG
CTTCTTCCTCAGATGATATATGGTTGGGT

ACTTTTCATGGTACGCCACCTCGTCCACGATATTGCCAAAGATTGGGTGGCGCT
CGTGCTTTTTATCCTCCTCCACCAAAAAGGACTCCTCCAGCCTATGGAAGAAAG
AGTCATCCACCTTAGCCATCTCATTACTAAAGATCTCCTGCAGGTAGCAGATCC
GATTCTTTCTGCGGGTATATCTGCGCCGTGCTGTTCTTTTGAGCCGCGTGGCTTC
GGCCGTCTCCCCGGAGTCGAACAGGAGGGCGCCAATGAGGTTCTTCTTTATGCT
GTGGCGATCGGTATTGCCCAGAACTTTGAATTTTTTGCTCGGCACCTTGTACTCG
TCCGTAATGACGGCCCAGCCGACGCTGTTTGTGCCGATATCGAGCCCAATGGAG
TACTTCTTGTCCATGGTGGCGGCTCTTGAAGGACGACGTCATCATCCCTTGCCC
GGATGCGCGGGCTTCTTGTCTAGCACAGGAGCCTGGGGTAGAGCGCATGCAAA
TTACGCGCTGTGCTTTGTGGGAAATCACCCTAAACGAAAAATTTATTCCTCTTTC
GAGCCTTATAGTGGCGGCCGGTCTACATCCTAGGTTTTAGAGCTAGAAATAGCA
AGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGT
GCTTTTTTGTTAACGCGGCCGCCTAGAGTCGACCTGCAGGCATGCAAGCTTGGC
GTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCA
CACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGT
GAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAA
CCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTT
TGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGT
TCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCA
CAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAA
GGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCC
CCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGA
CAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTC
CTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAG
CGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTT
CGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCC
TTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCA
CTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGC
TACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATT
TGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTC
TTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAA
GCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTT
CTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTC

ATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGT
TTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGC
TTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTG
CCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCC
CCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAG
CAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTA
TCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCG
CCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCA
CGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGA
GTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCG
ATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCA
CTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTG
AGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTT
GCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTG
CTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTG
TTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCT
TTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAG
GCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTC
ATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGA
GCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC
ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACA
TTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGT
GATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTG
TCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTG
TTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGA
GAGTGCACCATAGCGGCCGCGGTACCCACACAAAAAACCAACACACAGATGTA
ATGAAAATAAAGATATTTTATT
TCACACCTTCCTCTTCTTCTTGGGGTCAGC (SEQ ID NO: 62)

mSPA_HDR rd10(250)right

CTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAA
TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAA
AGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACT

GCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCA
ACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGC
TCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAG
CGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGAT
AACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTA
AAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATC
ACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAAC
CCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGC
TCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGG
GAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGG
TCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCT
GCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTAT
CGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGC
GGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAAC
AGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGG
TAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTG
CAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCT
TTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGG
TCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAA
GTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAAT
GCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGT
TGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGG
CCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATC
AGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTT
TATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTT
CGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGC
TACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGG
TTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGT
TAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATC
ACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAG
ATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCA
TTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGG
GATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACG

TTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGAT
GTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTT
TCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGC
AAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTT
TTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATT
TGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAA
AAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAA
ATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAA
ACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATG
CCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGG
GGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATAG
CGGCCGCGGTACCCACACAAAAAACCAACACACAGATGTAATGAAAATAAAG
ATATTTTATTTCACACCTTCCTCTTCTTCTTGGGGTCAGCCCTGCTGTCTCCACC
GAGCTGAGAGAGGTCGATTCTTGTTTCATAGAGCCCCGTAATTGACTGATGAAT
CAGTGTGGCGTCCAGGACCTCCTTTGTAGAGGTGTACCGCTTTCTGTCTATGGT
GGTGTCGAAGTACTTGAAGGCTGCAGGCGCGCCCAAGTTGGTCAGAGTAAACA
AGTGGATAATGTTTTCTGCCTGCTCCCTGATGGGCTTATCCCTGTGCTTATTGTA
AGCAGAAAGCACCTTATCGAGGTTAGCGTCGGCGAGGATCACTCTTTTGGAGA
ATTCGCTTATTTGCTCGATGATCTCATCAAGGTAGTGTTTGTGTTGTTCCACGAA
CAGCTGCTTCTGCTCATTATCTTCGGGAGACCCTTTGAGCTTTTCATAGTGGCTG
GCCAGATACAAGAAATTAACGTATTTAGAGGGCAGTGCCAGCTCGTTACCTTTC
TGCAGCTCGCCCGCACTAGCGAGCATTCGTTTCCGGCCGTTTTCAAGCTCAAAG
AGAGAGTACTTGGGAAGCTTAATGATGAGGTCTTTTTTGACCTCTTTATATCCTT
TCGCCTCGAGAAAGTCGATGGGGTTTTTTTCGAAGCTTGATCGCTCCATGATTG
TGATGCCCAGCAGTTCCTTGACGCTTTTGAGTTTTTTAGACTTCCCTTTCTCCAC
TTTGGCCACAACCAGTACACTGTAAGCGACTGTAGGAGAATCGAATCCGCCGT
ATTTCTTGGGGTCCCAATCTTTTT
TGCGTGCGATCAGCTTGTCGCTGTTCCTTTTCGGGAGGATACTTTCCTTGGAGA
AGCCTCCGGTCTGTACTTCGGTCTTTTTAACGATGTTCACCTGCGGCATGGACA
GGACCTTCCGGACTGTCGCGAAATCCCTACCCTTGTCCCACACGATTTCTCCTGT
TTCTCCGTTTGTTTCGATAAGTGGTCGCTTCCGAATCTCTCCATTGGCCAGTGTA
ATCTCGGTCTTGAAAAAATTCATAATATTGCTGTAA

AAGAAGTACTTAGCGGTGGCCTTGCCTATTTCCTGCTCAGACTTTGCGATCATTT
TCCTAACATCGTACACTTTATAGTCTCCGTAAACAAATTCAGATTCAAGCTTGG
GATATTTTTTGATAAGTGCAGTGCCTACCACTGCATTCAGGTAGGCATCATGCG
CATGGTGGTAATTGTTGATCTCTCTCACCTTATAAAACTGAAAGTCCTTTCTGAA
ATCTGAGACCAGCTTAGACTTCAGAGTAATAACTTTCACCTCTCGAATCAGTTT
GTCATTTTCATCGTACTTGGTGTTCATGCGTGAATCGAGAATTTGGGCCACGTG
CTTGGTGATCTGGCGTGTCTCAACAAGCTGCCTTTTGATGAAGCCGGCTTTATC
CAACTCAGACAGGCCACCTCGTTCAGCCTTAGTCAGATTATCGAACTTCCGTTG
TGTGATCAGTTTGGCGTTCAGCAGCTGCCGCCAATAATTTTTCATTTTCTTGACA
ACTTCTTCTGAGGGGACGTTATCACTCTTCCCTCTATTTTTATCGGATCTTGTCA
ACACTTTATTATCAATAGAATCATCTTTGAGAAAAGACTGGGGCACGATATGAT
CCACGTCGTAGTCGGAGAGCCGATTGATGTCCAGTTCCTGATCCACGTACATGT
CCCTGCCGTTCTGCAGGTAGTACAGGTAGAGCTTCTCATTCTGAAGCTGGGTGT
TTTCAACTGGGTGTTCCTTAAGGATTTGGGACCCCAGTTCTTTTATACCCTCTTC
AATCCTCTTCATCCTTTCCCTACTGTTCTTCTGTCCCTTCTGGGTAGTTTGGTTCT
CTCGGGCCATCTCGATAACGATATTCTCGGGCTTATGCCTTCCCATTACTTTGAC
GAGTTCATCCACGACCTTAACGGTCTGCAGTATTCCCTTTTTGATAGCTGGGCT
ACCTGCAAGATTAGCGATGTGCTCGTGAAGACTGTCCCCCTGGCCAGAAACTTG
TGCTTTCTGGATGTCCTCCTTAAAGGTGAGAGAGTCATCATGGATCAACTGCAT
GAAGTTCCGGTTGGCAAATCCATCGGACTTAAGAAAATCCAGGATTGTCTTTCC
ACTCTGCTTGTCTCGGATCCCATTGATCAGTTTTCTTGACAGCCGCCCCCATCCT
GTATATCGGCGCCTCTTGAGCTGTTTCATGACTTTGTCGTCGAAGAGATGAGCG
TAAGTTTTCAAGCGTTCTTCAATCATCTCCCTATCTTCAAACAACGTAAGGGTG
AGGACAATGTCCTCAAGAATGTCCTCGTTCTCCTCATTGTCCAGGAAGTCCTTG
TCTTTAATGATTTTCAGGAGATCGTGATACGTTCCCAGGGATGCGTTGAAGCGA
TCCTCCACTCCGCTGATTTCAACAGAGTCGAAACATTCAATCTTTTTGAAATAG
TCTTCTTTGAGCTGTTTCACGGTAACTTTCCGGTTCGTCTTGAAGAGGAGGTCCA
CGATAGCTTTCTTCTGCTCTCCAGACAGGAATGCTGGCTTTCTCATCCCTTCTGT
GACGTATTTGACCTTGGTGAGCTCGTTATAAACTGTGAAGTACTCGTACAGCAG
AGAGTGTTTAGGAAGCACCTTTTCGTTAGGCAGATTTTTATCAAAGTTAGTCAT
CCTTTCGATGAAGGACTGGGCAGAGGCCCCCTTATCCACGACTTCCTCGAAGTT
CCAGGGAGTGATGGTCTCTTCTGATTTGCGAGTCATCCACGCGAATCTGGAATT
TCCCCGGGCGAGGGGGCCTACATAGTAGGGTATCCGAAATGTGAGGATTTTCTC

AATCTTTTCCCTGTTATCTTTCAAAAAGGGGTAGAAATCCTCTTGCCGCCTGAG
GATAGCGTGCAGTTCGCCCAGGTGAATCTGGTGGGGGATGCTTCCATTGTCGAA
AGTGCGCTGTTTGCGCAACAGATCTTCTCTGTTAAGCTTTACCAGCAGCTCCTC
GGTGCCGTCCATTTTTTCCAAGATGGGCTTAATAAATTTGTAAAATTCCTCCTGG
CTTGCTCCGCCGTCAATGTATCCGGCGTAGCCATTTTTAGACTGATCGAAGAAA
ATTTCCTTGTACTTCTCAGGCAGTTGCTGTCTGACAAGGGCCTTCAGCAAAGTC
AAGTCTTGGTGGTGCTCATCATAGCGCTTGATCATACTAGCGCTCAGCGGAGCT
TTGGTGATCTCCGTGTTCACTCGCAGAATATCACTCAGCAGAATGGCGTCTGAC
AGGTTCTTTGCCGCCAAAAAAAGGTCTGCGTACTGGTCGCCGATCTGGGCCAGC
AGATTGTCGAGATCATCATCGTAGGTGTCTTTGCTCAGTTGAAGCTTGGCATCT
TCGGCCAGGTCGAAGTTAGATTTAAAGTTGGGGGTCAGCCCGAGTGACAGGGC
GATAAGATTACCAAACAGGCCGTTCTTCTTCTCCCCAGGGAGCTGTGCGATGAG
GTTTTCGAGCCGCCGGGATTTGGACAGCCTAGCGCTCAGGATTGCTTTGGCGTC
AACTCCGGATGCGTTGATCGGGTTCTCTTCGAAAAGCTGATTGTAAGTCTGAAC
CAGTTGGATAAAGAGTTTGTCGACATCGCTGTTGTCTGGGTTCAGGTCCCCCTC
GATGAGGAAGTGTCCCCGAAATTTGATCATATGCGCCAGCGCGAGATAGATCA
ACCGCAAGTCAGCCTTATCAGTACTGTCTACAAGCTTCTTCCTCAGATGATATA
TGGTTGGGTACTTTTCATGGTACGCCACCTCGTCCACGATATTGCCAAAGATTG
GGTGGCGCTCGTGCTTTTTATCCTCCTCCACCAAAAAGGACTCCTCCAGCCTAT
GGAAGAAAGAGTCATCCACCTTAGCCATCTCATTACTAAAGATCTCCTGCAGGT
AGCAGATCCGATTCTTTCTGCGGGTATATCTGCGCCGTGCTGTTCTTTTGAGCCG
CGTGGCTTCGGCCGTCTCCCCGGAGTCGAACAGGAGGGCGCCAATGAGGTTCTT
CTTTATGCTGTGGCGATCGGTATTGCCCAGAACTTTGAATTTTTTGCTCGGCACC
TTGTACTCGTCCGTAATGACGGCCCAGCCGACGCTGTTTGTGCCGATATCGAGC
CCAATGGAGTACTTCTTGTCCATGGTGGCGGCTCTTGAAGGACGACGTCATCAT
CCCTTGCCCGGATGCGCGGGCTTCTTGTCTAGCACAGGAGCCTGGGGTAGAGCG
CATGC
AAATTACGCGCTGTGCTTTGTGGGAAATCACCCTAAACGAAAAATTTATTCCTC
TTTCGAGCCTTATAGTGGCGGCCGGTCTACATCCTAGGTTTTAGAGCTAGAAAT
AGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT
CGGTGCTTTTTTGTTGATATGGTGCTGTGTAGGCTCATATGTGGATCTCAGAACC
CACATGTACTCTGCTCCCCAGGTCTTGGTGCGCTTTCTATTCTCTGTCAGCAAAG
CCTATCGAAGAATCACCTACCACAACTGGCGCCACGGCTTCAATGTAGCCCAG

ACCATGTTTACCCTACTCATGGTACGTATGTAAATTGGATGGGCTAGATGAATC
AGAGGGCTGGGGCAAGGACCACAGCTAACTATCTTCTGGCCCAAGGAACGCGG
CCGCCTAGAGTCGAC (SEQ ID NO: 63)

mSPA_HDR rd10(250)left

ATGTAATGAAAATAAAGATATTTTATTTCACACCTTCCTCTTCTTCTTGGGGTCA
GCCCTGCTGTCTCCACCGAGCTGAGAGAGGTCGATTCTTGTTTCATAGAGCCCC
GTAATTGACTGATGAATCAGTGTGGCGTCCAGGACCTCCTTTGTAGAGGTGTAC
CGCTTTCTGTCTATGGTGGTGTCGAAGTACTTGAAGGCTGCAGGCGCGCCCAAG
TTGGTCAGAGTAAACAAGTGGATAATGTTTTCTGCCTGCTCCCTGATGGGCTTA
TCCCTGTGCTTATTGTAAGCAGAAAGCACCTTATCGAGGTTAGCGTCGGCGAGG
ATCACTCTTTTGGAGAATTCGCTTATTTGCTCGATGATCTCATCAAGGTAGTGTT
TGTGTTGTTCCACGAACAGCTGCTTCTGCTCATTATCTTCGGGAGACCCTTTGAG
CTTTTCATAGTGGCTGGCCAGATACAAGAAATTAACGTATTTAGAGGGCAGTGC
CAGCTCGTTACCTTTCTGCAGCTCGCCCGCACTAGCGAGCATTCGTTTCCGGCC
GTTTTCAAGCTCAAAGAGAGAGTACTTGGGAAGCTTAATGATGAGGTCTTTTTT
GACCTCTTTATATCCTTTCGCCTCGAGAAAGTCGATGGGGTTTTTTTCGAAGCTT
GATCGCTCCATGATTGTGATGCCCAGCAGTTCCTTGACGCTTTTGAGTTTTTTAG
ACTTCCCTTTCTCCACTTTGGCCACAACCAGTACACTGTAAGCGACTGTAGGAG
AATCGAATCCGCCGTATTTCTTGGGGTCCCAATCTTTTTTGCGTGCGATCAGCTT
GTCGCTGTTCCTTTTCGGGAGGATACTTTCCTTGGAGAAGCCTCCGGTCTGTACT
TCGGTCTTTTTAACGATGTTCACCTGCGGCATGGACAGGACCTTCCGGACTGTC
GCGAAATCCCTACCCTTGTCCCACACGATTTCTCCTGTTTCTCCGTTTGTTTCGA
TAAGTGGTCGCTTCCGAATCTCTCCATTGGCCAGTGTAATCTCGGTCTTGAAAA
AATTCATAATATTGCTGTAAAAGAAGTACTTAGCGGTGGCCTTGCCTATTTCCT
GCTCAGACTTTGCGATCATTTTCCTAACATCGTACACTTTATAGTCTCCGTAAAC
AAATTCAGATTCAAGCTTGGGATATTTTTTGATAAGTGCAGTGCCTACCACTGC
ATTCAGGTAGGCATCATGCGCATGGTGGTAATTGTTGATCTCTCTCACCTTATA
AAACTGAAAGTCCTTTCTGAAATCTGAGACCAGCTTAGACTTCAGAGTAATAAC
TTTCACCTCTCGAATCAGTTTGTCATTTTCATCGTACTTGGTGTTCATGCGTGAA
TCGAGAATTTGGGCCACGTGCTTGGTGATCTGGCGTGTCTCAACAAGCTGCCTT
TTGATGAAGCCGGCTTTATCCAACTCAGACAGGCCACCTCGTTCAGCCTTAGTC
AGATTATCGAACTTCCGTTGTGTGATCAGTTTGGCGTTCAGCAGCTGCCGCCAA
TAATTTTTCATTTTCTTGACAACTTCTTCTGAGGGGACGTTATCACTCTTCCCTCT

ATTTTTATCGGATCTTGTCAACACTTTATTATCAATAGAATCATCTTTGAGAAAA
GACTGGGGCACGATATGATCCACGTCGTAGTCGGAGAGCCGATTGATGTCCAG
TTCCTGATCCACGTACATGTCCCTGCCGTTCTGCAGGTAGTACAGGTAGAGCTT
CTCATTCTGAAGCTGGGTGTTTTCAACTGGGTGTTCCTTAAGGATTTGGGACCC
CAGTTCTTTTATACCCTCTTCAATCCTCTTCATCCTTTCCCTACTGTTCTTCTGTC
CCTTCTGGGTAGTTTGGTTCTCTCGGGCCATCTCGATAACGATATTCTCGGGCTT
ATGCCTTCCCATTACTTTGACGAGTTCATCCACGACCTTAACGGTCTGCAGTATT
CCCTTTTTGATAGCTGGGCTACCTGCAAGATTAGCGATGTGCTCGTGAAGACTG
TCCCCCTGGCCAGAAACTTGTGCTTTCTGGATGTCCTCCTTAAAGGTGAGAGAG
TCATCATGGATCAACTGCATGAAGTTCCGGTTGGCAAATCCATCGGACTTAAGA
AAATCCAGGATTGTCTTTCCACTCTGCTTGTCTCGGATCCCATTGATCAGTTTTC
TTGACAGCCGCCCCCATCCTGTATATCGGCGCCTCTTGAGCTGTTTCATGACTTT
GTCGTCGAAGAGATGAGCGTAAGTTTTCAAGCGTTCTTCAATCATCTCCCTATC
TTCAAACAACGTAAGGGTGAGGACAATGTCCTCAAGAATGTCCTCGTTCTCCTC
ATTGTCCAGGAAGTCCTTGTCTTTAATGATTTTCAGGAGATCGTGATACGTTCCC
AGGGATGCGTTGAAGCGATCCTCCACTCCGCTGATTTCAACAGAGTCGAAACAT
TCAATCTTTTTGAAATAGTCTTCTTTGAGCTGTTTCACGGTAACTTTCCGGTTCG
TCTTGAAGAGGAGGTCCACGATAGCTTTCTTCTGCTCTCCAGACAGGAATGCTG
GCTTTCTCATCCCTTCTGTGACGTATTTGACCTTGGTGAGCTCGTTATAAACTGT
GAAGTACTCGTACAGCAGAGAGTGTTTAGGAAGCACCTTTTCGTTAGGCAGATT
TTTATCAAAGTTAGTCATCCTTTCGATGAAGGACTGGGCAGAGGCCCCCTTATC
CACGACTTCCTCGAAGTTCCAGGGAGTGATGGTCTCTTCTGATTTGCGAGTCAT
CCACGCGAATCTGGAATTTCCCCGGGCGAGGGGGCCTACATAGTAGGGTATCC
GAAATGTGAGGATTTTCTCAATCTTTTCCCTGTTATCTTTCAAAAAGGGGTAGA
AATCCTCTTGCCGCCTGAGGATAGCGTGCAGTTCGCCCAGGTGAATCTGGTGGG
GGATGCTTCCATTGTCGAAAGTGCGCTGTTTGCGCAACAGATCTTCTCTGTTAA
GCTTTACCAGCAGCTCCTCGGTGCCGTCCATTTTTTCCAAGATGGGCTTAATAA
ATTTGTAAAATTCCTCCTGGCTTGCTCCGCCGTCAATGTATCCGGCGTAGCCATT
TTTAGACTGATCGAAGAAAATTTCCTTGTACTTCTCAGGCAGTTGCTGTCTGAC
AAGGGCCTTCAGCAAAGTCAAGTCTTGGTGGTGCTCATCATAGCGCTTGATCAT
ACTAGCGCTCAGCGGAGCTTTGGTGATCTCCGTGTTCACTCGCAGAATATCACT
CAGCAGAATGGCGTCTGACAGGTTCTTTGCCGCCAAAAAAAGGTCTGCGTACT
GGTCGCCGATCTGGGCCAGCAGATTGTCGAGATCATCATCGTAGGTGTCTTTGC

TCAGTTGAAGCTTGGCATCTTCGGCCAGGTCGAAGTTAGATTTAAAGTTGGGGG
TCAGCCCGAGTGACAGGGCGATAAGATTACCAAACAGGCCGTTCTTCTTCTCCC
CAGGGAGCTGTGCGATGAGGTTTTCGAGCCGCCGGGATTTGGACAGCCTAGCG
CTCAGGATTGCTTTGGCGTCAACTCCGGATGCGTTGATCGGGTTCTCTTCGAAA
AGCTGATTGTAAGTCTGAACCAGTTGGATAAAGAGTTTGTCGACATCGCTGTTG
TCTGGGTTCAGGTCCCCCTCGATGAGGAAGTGTCCCCGAAATTTGATCATATGC
GCCAGCGCGAGATAGATCAACCGCAAGTCAGCCTTATCAGTACTGTCTACAAG
CTTCTTCCTCAGATGATATATGGTTGGGTACTTTTCATGGTACGCCACCTCGTCC
ACGATATTGCCAAAGATTGGGTGGCGCTCGTGCTTTTTATCCTCCTCCACCAAA
AAGGACTCCTCCAGCCTATGGAAGAAAGAGTCATCCACCTTAGCCATCTCATTA
CTAAAGATCTCCTGCAGGTAGCAGATCCGATTCTTTCTGCGGGTATATCTGCGC
CGTGCTGTTCTTTTGAGCCGCGTGGCTTCGGCCGTCTCCCCGGAGTCGAACAGG
AGGGCGCCAATGAGGTTCTTCTTTATGCTGTGGCGATCGGTATTGCCCAGAACT
TTGAATTTTTTGCTCGGCACCTTGTACTCGTCCGTAATGACGGCCCAGCCGACG
CTGTTTGTGCCGATATCGAGCCCAATGGAGTACTTCTTGTCCATGGTGGCGGCT
CTTGAAGGACGACGTCATCATCCCTTGCCCGGATGCGCGGGC
TTCTTGTCTAGCACAGGAGCCTGGGGTAGAGCGCATGCAAATTACGCGCTGTGC
TTTGTGGGAAATCACCCTAAACGAAAAATTTATTCCTCTTTCGAGCCTTATAGT
GGCGGCCGGTCTACATCCTAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAA
GGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTGTTA
ACGCGGCCGCCTAGAGTCGACCTGCAGGCATGCAAGC
TTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAA
TTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAA
TGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCG
GGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGG
CGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCT
CACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTC
AAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAAC
ATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGC
TGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCT
CAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAA
GATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCT
GCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCT

CATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTG
GGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAAC
TATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCC
ACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTT
GAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCG
CTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCA
AACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGC
GCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACG
CTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAA
AGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAA
AGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCA
CCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCG
T
GTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGA
TACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCA
GCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCA
GTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTT
GCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGG
TATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCC
CATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAG
TAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCT
TACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAA
GTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAAT
ACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAA
AACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTT
CGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAG
CGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATA
AGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGA
AGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAG
AAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGA
CGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCAC
GAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACAT
GCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGAC

AAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAAC
TATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATAGCGGCCGCGGTAC
GATATGGTGCTGTGTAGGCTCATATGTGGATCTCAGAACCCACATGTACTCTGC
TCCCCAGGTCTTGGTGCGCTTTCTATTCTCTGTCAGCAAAGCCTATCGAAGAAT
CACCTACCACAACTGGCGCCACGGCTTCAATGTAGCCCAGACCATGTTTACCCT
ACTCATGGTACGTATGTAAATTGGATGGGCTAGATGAATCAGAGGGCTGGGGC
AAGGACCACAGCTAACTATCTTCTGGCCCAAGGGTACCCACACAAAAAACCAA
CACACAG (SEQ ID NO: 64)

mSPA_HDR rd12(250)right

CTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAA
TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAA
AGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACT
GCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCA
ACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCAC
TGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAA
GGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGT
GAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGC
GTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAG
TCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTG
GAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTC
CGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTAT
CTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCC
GTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCG
GTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAG
AGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACG
GCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCT
TCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGC
GGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAA
GAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCA
CGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTT
TTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGG
TCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTA
TTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGG

GAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTC
ACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCA
GAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGG
AAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTG
CTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCG
GTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCG
GTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTA
TCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAA
GATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTA
TGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCAC
ATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAA
CTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCA
CCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAA
ACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTT
GAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTG
TCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGG
TTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTA
TCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGC
GTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTC
ACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTC
AGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGA
TTGTACTGAGAGTGCACCATAGCGGCCGCGGTACCCACACAAAAAACCAACAC
ACAGATGTAATGAAAATAAAGATATTTTATTTCACACCTTCCTCTTCTTCTTGGG
GTCAGCCCTGCTGTCTCCACCGAGCTGAGAGAGGTCGATTCTTGTTTCATAGAG
CCCCGTAATTGACTGATGAATCAGTGTGGCGTCCAGGACCTCCTTTGTAGAGGT
GTACCGCTTTCTGTCTATGGTGGTGTCGAAGTACTTGAAGGCTGCAGGCGCGCC
CAAGTTGGTCAGAGTAAACAAGTGGATAATGTTTTCTGCCTGCTCCCTGATGGG
CTTATCCCTGTGCTTATTGTAAGCAGAAAGCACCTTATCGAGGTTAGCGTCGGC
GAGGATCACTCTTTTGGAGAATTCGCTTATTTGCTCGATGATCTCATCAAGGTA
GTGTTTGTGTTGTTCCACGAACAGCTGCTTCTGCTCATTATCTTCGGGAGACCCT
TTGAGCTTTTCATAGTGGCTGGCCAGATACAAGAAATTAACGTATTTAGAGGGC
AGTGCCAGCTCGTTACCTTTCTGCAGCTCGCCCGCACTAGCGAGCATTCGTTTC
CGGCCGTTTTCAAGCTCAAAGAGAGAGTACTTGGGAAGCTTAATGATGAGGTC

TTTTTTGACCTCTTTATATCCTTTCGCCTCGAGAAAGTCGATGGGGTTTTTTTCG
AAGCTTGATCGCTCCATGATTGTGATGCCCAGCAGTTCCTTGACGCTTTTGAGTT
TTTTAGACTTCCCTTTCTCCACTTTGGCCACAACCAGTACACTGTAAGCGACTGT
AGGAGAATCGAATCCGCCGTATTTCTTGGGGTCCCAATCTTTTTTGCGTGCGAT
CAGCTTGTCGCTGTTCCTTTTCGGGAGGATACTTTCCTTGGAGAAGCCTCCGGTC
TGTACTTCGGTCTTTTTAACGATGTTCACCTGCGGCATGGACAGGACCTTCCGG
ACTGTCGCGAAATCCCTACCCTTGTCCCACACGATTTCTCCTGTTTCTCCGTTTG
TTTCGATAAGTGGTCGCTTCCGAATCTCTCCATTGGCCAGTGTAATCTCGGTCTT
GAAAAAATTCATAATATTGCTGTAAAAGAAGTACTTAGCGGTGGCCTTGCCTAT
TTCCTGCTCAGACTTTGCGATCATTTTCCTAACATCGTACACTTTATAGTCTCCG
TAAACAAATTCAGATTCAAGCTTGGGATATTTTTTGATAAGTGCAGTGCCTACC
ACTGCATTCAGGTAGGCATCATGCGCATGGTGGTAATTGTTGATCTCTCTCACC
TTATAAAACTGAAAGTCCTTTCTGAAATCTGAGACCAGCTTAGACTTCAGAGTA
ATAACTTTCACCTCTCGAATCAGTTTGTCATTTTCATCGTACTTGGTGTTCATGC
GTGAATCGAGAATTTGGGCCACGTGCTTGGTGATCTGGCGTGTCTCAACAAGCT
GCCTTTTGATGAAGCCGGCTTTATCCAACTCAGACAGGCCACCTCGTTCAGCCT
TAGTCAGATTATCGAACTTCCGTTGTGTGATCAGTTTGGCGTTCAGCAGCTGCC
GCCAATAATTTTTCATTTTCTTGACAACTTCTTCTGAGGGGACGTTATCACTCTT
CCCTCTATTTTTATCGGATCTTGTCAACACTTTATTATCAATAGAATCATCTTTG
AGAAAAGACTGGGGCACGATATGATCCACGTCGTAGTCGGAGAGCCGATTGAT
GTCCAGTTCCTGATCCACGTACATGTCCCTGCCGTTCTGCAGGTAGTACAGGTA
GAGCTTCTCATTCTGAAGCTGGGTGTTTTCAACTGGGTGTTCCTTAAGGATTTGG
GACCCCAGTTCTTTTATACCCTCTTCAATCCTCTTCATCCTTTCCCTACTGTTCTT
CTGTCCCTTCTGGGTAGTTTGGTTCTCTCGGGCCATCTCGATAACGATATTCTCG
GGCTTATGCCTTCCCATTACTTTGACGAGTTCATCCACGACCTTAACGGTCTGCA
GTATTCCCTTTTTGATAGCTGGGCTACCTGCAAGATTAGCGATGTGCTCGTGAA
GACTGTCCCCCTGGCCAGAAACTTGTGCTTTCTGGATGTCCTCCTTAAAGGTGA
GAGAGTCATCATGGATCAACTGCATGAAGTTCCGGTTGGCAAATCCATCGGACT
TAAGAAAATCCAGGATTGTCTTTCCACTCTGCTTGTCTCGGATCCCATTGATCA
GTTTTCTTGACAGCCGCCCCCATCCTGTATATCGGCGCCTCTTGAGCTGTTTCAT
GACTTTGTCGTCGAAGAGATGAGCGTAAGTTTTCAAGCGTTCTTCAATCATCTC
CCTATCTTCAAACAACGTAAGGGTGAGGACAATGTCCTCAAGAATGTCCTCGTT
CTCCTCATTGTCCAGGAAGTCCTTGTCTTTAATGATTTTCAGGAGATCGTGATAC

GTTCCCAGGGATGCGTTGAAGCGATCCTCCACTCCGCTGATTTCAACAGAGTCG
AAACATTCAATCTTTTTGAAATAGTCTTCTTTGAGCTGTTTCACGGTAACTTTCC
GGTTCGTCTTGAAGAGGAGGTCCACGATAGCTTTCTTCTGCTCTCCAGACAGGA
ATGCTGGCTTTCTCATCCCTTCTGTGACGTATTTGACCTTGGTGAGCTCGTTATA
AACTGTGAAGTACTCGTACAGCAGAGAGTGTTTAGGAAGCACCTTTTCGTTAGG
CAGATTTTTATCAAAGTTAGTCATCCTTTCGATGAAGGACTGGGCAGAGGCCCC
CTTATCCACGACTTCCTCGAAGTTCCAGGGAGTGATGGTCTCTTCTGATTTGCG
AGTCATCCACGCGAATCTGGAATTTCCCCGGGCGAGGGGGCCTACATAGTAGG
GTATCCGAAATGTGAGGATTTTCTCAATCTTTTCCCTGTTATCTTTCAAAAAGGG
GTAGAAATCCTCTTGCCGCCTGAGGATAGCGTGCAGTTCGCCCAGGTGAATCTG
GTGGGGGATGCTTCCATTGTCGAAAGTGCGCTGTTTGCGCAACAGATCTTCTCT
GTTAAGCTTTACCAGCAGCTCCTCGGTGCCGTCCATTTTTTCCAAGATGGGCTTA
ATAAATTTGTAAAATTCCTCCTGGCTTGCTCCGCCGTCAATGTATCCGGCGTAG
CCATTTTTAGACTGATCGAAGAAAATTTCCTTGTACTTCTCAGGCAGTTGCTGTC
TGACAAGGGCCTTCAGCAAAGTCAAGTCTTGGTGGTGCTCATCATAGCGCTTGA
TCATACTAGCGCTCAGCGGAGCTTTGGTGATCTCCGTGTTCACTCGCAGAATAT
CACTCAGCAGAATGGCGTCTGACAGGTTCTTTGCCGCCAAAAAAAGGTCTGCGT
ACTGGTCGCCGATCTGGGCCAGCAGATTGTCGAGATCATCATCGTAGGTGTCTT
TGCTCAGTTGAAGCTTGGCATCTTCGGCCAGGTCGAAGTTAGATTTAAAGTTGG
GGGTCAGCCCGAGTGACAGGGCGATAAGATTACCAAACAGGCCGTTCTTCTTCT
CCCCAGGGAGCTGTGCGATGAGGTTTTCGAGC
CGCCGGGATTTGGACAGCCTAGCGCTCAGGATTGCTTTGGCGTCAACTCCGGAT
GCGTTGATCGGGTTCTCTTCGAAAAGCTGATTGTAAGTCTGAACCAGTTGGATA
AAGAGTTTGTCGACATCGCTGTTGTCTGGGTTCAGGTCCCCCTCGATGAGGAAG
TGTCCCCGAAATTTGATCATATGCGCCAGCGCGAGATAGATCAACCGCAAGTC
AGCCTTATCAGTACTGTCTACAAGCTTCTTCCTCAG
ATGATATATGGTTGGGTACTTTTCATGGTACGCCACCTCGTCCACGATATTGCC
AAAGATTGGGTGGCGCTCGTGCTTTTTATCCTCCTCCACCAAAAAGGACTCCTC
CAGCCTATGGAAGAAAGAGTCATCCACCTTAGCCATCTCATTACTAAAGATCTC
CTGCAGGTAGCAGATCCGATTCTTTCTGCGGGTATATCTGCGCCGTGCTGTTCTT
TTGAGCCGCGTGGCTTCGGCCGTCTCCCCGGAGTCGAACAGGAGGGCGCCAAT
GAGGTTCTTCTTTATGCTGTGGCGATCGGTATTGCCCAGAACTTTGAATTTTTTG
CTCGGCACCTTGTACTCGTCCGTAATGACGGCCCAGCCGACGCTGTTTGTGCCG

ATATCGAGCCCAATGGAGTACTTCTTGTCCATGGTGGCGGCTCTTGAAGGACGA
CGTCATCATCCCTTGCCCGGATGCGCGGGCTTCTTGTCTAGCACAGGAGCCTGG
GGTAGAGCGCATGCAAATTACGCGCTGTGCTTTGTGGGAAATCACCCTAAACG
AAAAATTTATTCCTCTTTCGAGCCTTATAGTGGCGGCCGGTCTACATCCTAGGTT
TTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAA
AAGTGGCACCGAGTCGGTGCTTTTTTGTTGTTCTGTCTATAGGTAAGCTGACAA
ATAACAAATAGGCACATAGAAAATCTAGTAAGTAGTACCACCTGATATCTCAC
TTTGCTGCAGGCAGGATTCCCCTCTGGCTCACTGGCAGTCTCCTCCGGTGTGGG
CCAGGGCTCTTTGAAGTTGGATCTGAGCCTTTCTATCACCTGTTTGATGGACAA
GCCCTTTTGCACAAGTTTGACTTCAAGGAGGGCCATGTCACATACCACAGAAGG
TAAGTCCATGAACGCGGCCGCCTAGAGTCGAC (SEQ ID NO: 65)

mSPA_HDR rd12(250)left

ATGTAATGAAAATAAAGATATTTTATTTCACACCTTCCTCTTCTTCTTGGGGTCA
GCCCTGCTGTCTCCACCGAGCTGAGAGAGGTCGATTCTTGTTTCATAGAGCCCC
GTAATTGACTGATGAATCAGTGTGGCGTCCAGGACCTCCTTTGTAGAGGTGTAC
CGCTTTCTGTCTATGGTGGTGTCGAAGTACTTGAAGGCTGCAGGCGCGCCCAAG
TTGGTCAGAGTAAACAAGTGGATAATGTTTTCTGCCTGCTCCCTGATGGGCTTA
TCCCTGTGCTTATTGTAAGCAGAAAGCACCTTATCGAGGTTAGCGTCGGCGAGG
ATCACTCTTTTGGAGAATTCGCTTATTTGCTCGATGATCTCATCAAGGTAGTGTT
TGTGTTGTTCCACGAACAGCTGCTTCTGCTCATTATCTTCGGGAGACCCTTTGAG
CTTTTCATAGTGGCTGGCCAGATACAAGAAATTAACGTATTTAGAGGGCAGTGC
CAGCTCGTTACCTTTCTGCAGCTCGCCCGCACTAGCGAGCATTCGTTTCCGGCC
GTTTTCAAGCTCAAAGAGAGAGTACTTGGGAAGCTTAATGATGAGGTCTTTTTT
GACCTCTTTATATCCTTTCGCCTCGAGAAAGTCGATGGGGTTTTTTTCGAAGCTT
GATCGCTCCATGATTGTGATGCCCAGCAGTTCCTTGACGCTTTTGAGTTTTTTAG
ACTTCCCTTTCTCCACTTTGGCCACAACCAGTACACTGTAAGCGACTGTAGGAG
AATCGAATCCGCCGTATTTCTTGGGGTCCCAATCTTTTTTGCGTGCGATCAGCTT
GTCGCTGTTCCTTTTCGGGAGGATACTTTCCTTGGAGAAGCCTCCGGTCTGTACT
TCGGTCTTTTTAACGATGTTCACCTGCGGCATGGACAGGACCTTCCGGACTGTC
GCGAAATCCCTACCCTTGTCCCACACGATTTCTCCTGTTTCTCCGTTTGTTTCGA
TAAGTGGTCGCTTCCGAATCTCTCCATTGGCCAGTGTAATCTCGGTCTTGAAAA
AATTCATAATATTGCTGTAAAAGAAGTACTTAGCGGTGGCCTTGCCTATTTCCT
GCTCAGACTTTGCGATCATTTTCCTAACATCGTACACTTTATAGTCTCCGTAAAC

AAATTCAGATTCAAGCTTGGGATATTTTTTGATAAGTGCAGTGCCTACCACTGC
ATTCAGGTAGGCATCATGCGCATGGTGGTAATTGTTGATCTCTCTCACCTTATA
AAACTGAAAGTCCTTTCTGAAATCTGAGACCAGCTTAGACTTCAGAGTAATAAC
TTTCACCTCTCGAATCAGTTTGTCATTTTCATCGTACTTGGTGTTCATGCGTGAA
TCGAGAATTTGGGCCACGTGCTTGGTGATCTGGCGTGTCTCAACAAGCTGCCTT
TTGATGAAGCCGGCTTTATCCAACTCAGACAGGCCACCTCGTTCAGCCTTAGTC
AGATTATCGAACTTCCGTTGTGTGATCAGTTTGGCGTTCAGCAGCTGCCGCCAA
TAATTTTTCATTTTCTTGACAACTTCTTCTGAGGGGACGTTATCACTCTTCCCTCT
ATTTTTATCGGATCTTGTCAACACTTTATTATCAATAGAATCATCTTTGAGAAAA
GACTGGGGCACGATATGATCCACGTCGTAGTCGGAGAGCCGATTGATGTCCAG
TTCCTGATCCACGTACATGTCCCTGCCGTTCTGCAGGTAGTACAGGTAGAGCTT
CTCATTCTGAAGCTGGGTGTTTTCAACTGGGTGTTCCTTAAGGATTTGGGACCC
CAGTTCTTTTATACCCTCTTCAATCCTCTTCATCCTTTCCCTACTGTTCTTCTGTC
CCTTCTGGGTAGTTTGGTTCTCTCGGGCCATCTCGATAACGATATTCTCGGGCTT
ATGCCTTCCCATTACTTTGACGAGTTCATCCACGACCTTAACGGTCTGCAGTATT
CCCTTTTTGATAGCTGGGCTACCTGCAAGATTAGCGATGTGCTCGTGAAGACTG
TCCCCCTGGCCAGAAACTTGTGCTTTCTGGATGTCCTCCTTAAAGGTGAGAGAG
TCATCATGGATCAACTGCATGAAGTTCCGGTTGGCAAATCCATCGGACTTAAGA
AAATCCAGGATTGTCTTTCCACTCTGCTTGTCTCGGATCCCATTGATCAGTTTTC
TTGACAGCCGCCCCCATCCTGTATATCGGCGCCTCTTGAGCTGTTTCATGACTTT
GTCGTCGAAGAGATGAGCGTAAGTTTTCAAGCGTTCTTCAATCATCTCCCTATC
TTCAAACAACGTAAGGGTGAGGACAATGTCCTCAAGAATGTCCTCGTTCTCCTC
ATTGTCCAGGAAGTCCTTGTCTTTAATGATTTTCAGGAGATCGTGATACGTTCCC
AGGGATGCGTTGAAGCGATCCTCCACTCCGCTGATTTCAACAGAGTCGAAACAT
TCAATCTTTTTGAAATAGTCTTCTTTGAGCTGTTTCACGGTAACTTTCCGGTTCG
TCTTGAAGAGGAGGTCCACGATAGCTTTCTTCTGCTCTCCAGACAGGAATGCTG
GCTTTCTCATCCCTTCTGTGACGTATTTGACCTTGGTGAGCTCGTTATAAACTGT
GAAGTACTCGTACAGCAGAGAGTGTTTAGGAAGCACCTTTTCGTTAGGCAGATT
TTTATCAAAGTTAGTCATCCTTTCGATGAAGGACTGGGCAGAGGCCCCCTTATC
CACGACTTCCTCGAAGTTCCAGGGAGTGATGGTCTCTTCTGATTTGCGAGTCAT
CCACGCGAATCTGGAATTTCCCCGGGCGAGGGGGCCTACATAGTAGGGTATCC
GAAATGTGAGGATTTTCTCAATCTTTTCCCTGTTATCTTTCAAAAAGGGGTAGA
AATCCTCTTGCCGCCTGAGGATAGCGTGCAGTTCGCCCAGGTGAATCTGGTGGG

GGATGCTTCCATTGTCGAAAGTGCGCTGTTTGCGCAACAGATCTTCTCTGTTAA
GCTTTACCAGCAGCTCCTCGGTGCCGTCCATTTTTTCCAAGATGGGCTTAATAA
ATTTGTAAAATTCCTCCTGGCTTGCTCCGCCGTCAATGTATCCGGCGTAGCCATT
TTTAGACTGATCGAAGAAAATTTCCTTGTACTTCTCAGGCAGTTGCTGTCTGAC
AAGGGCCTTCAGCAAAGTCAAGTCTTGGTGGTGCTCATCATAGCGCTTGATCAT
ACTAGCGCTCAGCGGAGCTTTGGTGATCTCCGTGTTCACTCGCAGAATATCACT
CAGCAGAATGGCGTCTGACAGGTTCTTTGCCGCCAAAAAAAGGTCTGCGTACT
GGTCGCCGATCTGGGCCAGCAGATTGTCGAGATCATCATCGTAGGTGTCTTTGC
TCAGTTGAAGCTTGGCATCTTCGGCCAGGTCGAAGTTAGATTTAAAGTTGGGGG
TCAGCCCGAGTGACAGGGCGATAAGATTACCAAACAGGCCGTTCTTCTTCTCCC
CAGGGAGCTGTGCGATGAGGTTTTCGAGCCGCCGGGATTTGGACAGCCTAGCG
CTCAGGATTGCTTTGGCGTCAACTCCGGATGCGTTGATCGGGTTCTCTTCGAAA
AGCTGATTGTAAGTCTGAACCAGTTGGATAAAGAGTTTGTCGACATCGCTGTTG
TCTGGGTTCAGGTCCCCCTCGATGAGGAAGTGTCCCCGAAATTTGATCATATGC
GCCAGCGCGAGATAGATCAACCGCAAGTCAGCCTTATCAGTACTGTCTACAAG
CTTCTTCCTCAGATGATATATGGTTGGGTACTTTTCATGGTACGCCACCTCGTCC
ACGATATTGCCAAAGATTGGGTGGCGCTCGTGCTTTTTATCCTCCTCCACCAAA
AAGGACTCCTCCAGCCTATGGAAGAAAGAGTCATCCACCTTAGCCATCTCATTA
CTAAAGATCTCCTGCAGGTAGCAGATCCGATTCTTTCTGCGGGTATATCTGCGC
CGTGCTGTTCTTTTGAGCCGCGTGGCTTCGGCCGTCTCCCCGGAGTCGAACAGG
AGGGCGCCAATGAGGTTCTTCTTTATGCTGTGGCGATCGGTATTGCCCAGAACT
TTGAATTTTTTGCTCGGCACCTTGTACTCGTCCGTAATGACGGCCCAGCCGACG
CTGTTTGTGCCGATATCGAGCCCAATGGAGTACTTCTTGTCCATGGTGGCGGCT
CTTGAAGGACGACGTCATCATCCCTTGCCCGGATGCGCGGGC
TTCTTGTCTAGCACAGGAGCCTGGGGTAGAGCGCATGCAAATTACGCGCTGTGC
TTTGTGGGAAATCACCCTAAACGAAAAATTTATTCCTCTTTCGAGCCTTATAGT
GGCGGCCGGTCTACATCCTAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAA
GGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTGTTA
ACGCGGCCGCCTAGAGTCGACCTGCAGGCATGCAAGC
TTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAA
TTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAA
TGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCG

GGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGG
CGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCT
CACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTC
AAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAAC
ATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGC
TGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCT
CAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAA
GATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCT
GCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCT
CATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTG
GGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAAC
TATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCC
ACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTT
GAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCG
CTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCA
AACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGC
GCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACG
CTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAA
AGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAA
AGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCA
CCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCG
T
GTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGA
TACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCA
GCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCA
GTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTT
GCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGG
TATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCC
CATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAG
TAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCT
TACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAA
GTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAAT
ACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAA

AACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTT
CGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAG
CGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATA
AGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGA
AGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAG
AAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGA
CGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCAC
GAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACAT
GCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGAC
AAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAAC
TATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATAGCGGCCGCGGTTC
TGTCTATAGGTAAGCTGACAAATAACAAATAGGCACATAGAAAATCTAGTAAG
TAGTACCACCTGATATCTCACTTTGCT
GCAGGCAGGATTCCCCTCTGGCTCACTGGCAGTCTCCTCCGGTGTGGGCCAGGG
CTCTTTGAAGTTGGATCTGAGCCTTTCTATCACCTGTTTGATGGACAAGCCCTTT
TGCACAAGTTTGACTTCAAGGAGGGCCATGTCACATACCACAGAAGGTAAGTC
CATGGTACCCACACAAAAAACCAACACACAG (SEQ ID NO: 66)

mSPA_HDR SaCas9 entry plasmid

GCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCAC
AATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCT
AATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGT
CGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGA
GGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCT
CGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGG
TTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCC
AGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGG
CTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCG
AAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGT
GCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCT
TCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTG
TAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGAC
CGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGAC
TTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGT

AGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAA
GAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAG
TTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTG
TTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTG
ATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATT
TTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAA
TGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTAC
CAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCA
TAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCAT
CTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATT
TATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCA
ACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGT
AGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTG
GTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCA
AGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGT
CCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATG
GCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGA
CTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTT
GCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAA
AAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTAC
CGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAG
CATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATG
CCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTC
CTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACA
TATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCC
GAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATA
AAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTG
AAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCG
GATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTG
TCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACC
ATAGCGGCCGCGGTACCCACACAAAAAACCAACACACAGATGTAATGAAAATA
AAGATATTTTATTAGGCTGATCAGCGAGCTCTAGGAATTCTTAGGATCCCTTTTT
CTTTTTTGCCTGGCCGGCCTTTTTCGTGGCCGCCGGCCTTTTGCCCTTTTTGATG

ATCTGAGGGTGCTTCTTAGATTTCACTTCATACAGGTTGCCCAGAATGTCTGTG
CTGTACTTCTTAATGCTCTGGGTCTTGGAGGCGATTGTCTTAATGATCCTGGGG
GGCCTCTTGTCGTTCATGTTTTCCAGGTACTCGCGGTAGGTGATGTCGATCATGT
TCACTTCGATCCGGTTCAGCAGGTCGTTGTTCACGCCGATCACTCTATACAGCT
CGCCGTTGATCTTGATCAGATCGTTGTTGTAGAAGGAGGCGATAAACTCGGCCT
GGTTGCTGATCTTCTTCAGCTTCTTAGCTTCCTCATAGCACTTGCTATTCACTTC
GTAGTAGTTTTCTTTTTTGATCACATCCAGATTCTTCACGGTCACGAACTTGTAC
ACGCCATTGTCCAGGTACACGTCGAATCTGTAGGGCTTCAGGGACAGCTTCACG
ACCTTGTTTCTGCTGTTGGGGTAGTCGTCGGTGATGTCCAGATGGGCGTTCAGT
TTGTTGCCGTAATACTTAATCTTCTTGATCACGGGGCCGTTGTCCTTTTTGGAGT
ACTTGGTCAGGTAGTTCCCGGTTTCCTCGTAGTACTTGTACAGGGGATTCTTCTC
GTCGCCGTACTGTTCCATAATCAGCTTCAGTTTCTGGTAGGTCTGGGGGTCGTG
GTGGTACATCAGCAGCTTTTCGGGGCTCTTGTTGATCAGCTTTTTCAGCTTGTCA
TTGTCCTTGTCGTACAGGCCGTTCAGATTGTTCACGATCAGGGTGTTGCCCTTGT
CGTCCTTCCGGGTGGAGTACAGGGTGTCGTTAATCAGCTCTCTATTAGGCTTCTT
GTCCACCCGGTGGCTGTACTTGTAGTCCTTGAAGTCCTTAATGTGCTTGATCTGG
TGGGGGGTGATGAAGATCTCTTTGTACTCCTGCTCGGTTTCGATCTCGGGCATG
CTCTCGGCCTGCTTTTCCTCGAACATCTGGTTTTCCATCACTTTTTTGGCCTTGTC
CAGTTTCTTCCACTCTTTGAAGATGAAATCGGCGTTGGCAATGATCAGGGCGTC
CTCGGCGTGGTGCTTGTACCCCTTGTTCCGCTCTTTCTTAAACTTCCACTTCCGC
CGCAGAAAGCTGGTGAAGCCGCCATTGATGGACTTCACTTTCACGTCCAGGTTG
TTCACTCTGAAGTAGCTCCGCAGCAGGTTCATCAGGCCTCTGGTGGCGTATCTG
GTATCCACCAGGTTCCGGTTGATGAAGTCTTTCTGCACGGAGAACCTGTTGATG
TCCCGTTCTTCCAGCAGATACTCTTTCTTGGTCTTGCTGATTCTGCCCTTGCCCTT
GGCCAGATTCAGGATGTGCTTCTTGAAGGTTTCGTAGCTGATCTTGCTGTCGCT
GCTGCTCAGGTACTGGAATGGGGTCCGGTTGCCCTTCTTGCTGTTTTCTTCCTGC
TTCACGAGCACCTTGTTGTTGAAGCTGTTGTCGAAGGACACGCTTCTGGGGATG
ATGTGGTCCACCTCATAGTTGAAGGGGTTGTTCAGCAGATCTTCCAGAGGGATG
GCTTCCAGGCTGTACAGGCACTTGCCTTCCTGCATGTCGTGCAGCTTGATCTTCT
CGATCAGGTACTTGGCGTTCTCTTTGCCGGTGGTCCGGATGATTTCCTCGATCCG
CTCGTTGGTCTGCCGGTTCCGCTTCTGCATCTCGTTGATCATTTTCTGGGCGTCC
TTGGAGTTCTTCTCGCGGGCCAGCTCGATAATGATGTCGTTGGGCAGGCCGTAC
TTCTTGATGATGGCGTTGATCACTTTGATGCTCTGGATGAAGCTTCTCTTCACGA

CGGGGCTCAGGATGAAGTCGTCCACCAGGGTGGTGGGGATCTCTTTCTGCTGGG
ACAGGTCCACCTTCTTGGGCACCAGCTTCAGCCGGTTGAAGATAGCGATCTGGT
TGTCGTTGGTGTGCCACAGCTCGTCCAGGATCAGGTTGATGGCCTTCAGGCTCA
GGTTGTGGGTGCCGGTATAGCCCTTCAGATTAGAGATCTGCTCGATCTCTTCCT
GGGTCAGCTCGGAGTTCAGATTGGTCAGTTCTTCCTGGATGTCCTCGCTGCTCT
GGTAGATGGTCAGGATCTTGGCAATCTGATCCAGCAGCTCGGCGTTCTCAATAA
TCTCTTTCCGGGCGGTAATGTCCTTGATGTCGTGGTACACCTTCAGGTTGGTGA
ACTCGGGCTTGCCGGTGCTGGTCACTCTGTAGCCCTTAATATCCTCTTCGTTCAC
GAGGATTTCTTTGGCGATCTGCTTCAGGGTGGGCTTCTTCTTCTGCTTGAACACG
TTCTCGATGATCTGGAACTTCTCGTAATATTCCAGCTTCTCGTTCTCGTCCCTGG
TGATCACGAGATTGTTCAGGTCGTTCAGGGCGTTGTACAGGTCGGCGTTGTAGG
CGTACTTCACGCTCCGCAGTTCCTCGGGGAAGTAGGTGCAGTGGCCCATCAGCA
TCTCGTACCATTCTTTGATGTCCTTCCAGCCGAAGGGGCTGCCCTCGCCAGGTC
CCTCATAGTAGGTCCGCCGGGTTTCCAGCAGGTCGATGTAGGTGTCGATGAAGC
TCTGGTCCAGCTGGTGGTAGGCCTTCTGCACCTTCAGCAGCTGTTTGGCTTCTTT
CACGTAGTCGCTGGTCTTGAATCTGTTGATGCTGCCCCGCACTTCGCCGTCTTTC
TTCAGCCGTTCCAGCTGCAGTTCGGCCACGTATTTCTCTTCCAGGGCCTTGCTGT
TCCGGCTGATCTGCTCTTTGGTGGACAGCTCGTTGCCGGTGTCCTCTTCCACCTC
GTTCACGTTGT

GCACGCCTCTTCTCTTGGCCAGGTGCAGCAGGGCGGCAGAGAACTCTTCCTCGC
TCAGCTTCTGGCTCAGGCCCTTCACTCTGGCCTCGTAGGGGTTGATGCCGCTCA
GCTCGCTGTGGTCGGTCAGCAGGTTGTAGTCGAACAGCAGCTTCTTCACTCTCT
GGATTCTATGCCGCCTCCGCCGCTTCAGCCTTCTGGCGCCTCTCTTGCTCCGCCT
GCCCTCGTTGTTTTCCACGTTGGCCTCTTTGAACAGCCGCACGCCGGCATCGAT
CACGTCCCGTGTCTCGTAGTCGATGATGCCGTAGCCCACGCTGGTGATGCCGAT
GTCCAGGCCCAGGATGTAGTTCCGCTTGGCTGCTGGGACTCCGTGGATACCGAC
CTTCCGCTTCTTCTTTGGGGCCATGTGGCGGCTCTTGAAGGACGACGTCATCAT
CCCTTGCCCGGATGCGCGGGCTTCTTGTCTAGCACAGGAGCCTGGGGTAGAGCG
CATGCAAATTACGCGCTGTGCTTTGTGGGAAATCACCCTAAACGAAAAATTTAT
TCCTCTTTCGAGCCTTATAGTGGCGGCCGGTCTACATCCTAGGTTTTAGTACTCT
GGAAACAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTT
GTTGGCGAGATTTTTTGTTAACGCGGCCGCCTAGAGTCGACCTGCAGGCATGCA
A (SEQ ID NO: 67)

REFERENCES

All publications, patent applications, patents, and other references mentioned in the specification are indicative of the level of those skilled in the art to which the presently disclosed subject matter pertains. All publications, patent applications, patents, and other references are herein incorporated by reference to the same extent as if each individual publication, patent application, patent, and other reference was specifically and individually indicated to be incorporated by reference. It will be understood that, although a number of patent applications, patents, and other references are referred to herein, such reference does not constitute an admission that any of these documents forms part of the common general knowledge in the art.

Although the foregoing subject matter has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be understood by those skilled in the art that certain changes and modifications can be practiced within the scope of the appended claims.

CLAIMS

1. A non-naturally occurring nuclease system comprising one or more vectors comprising:

a bidirectional promoter selected from the group consisting of 7sk, SRP-RPS29, SNAR-G-l-CGBl, SNAR-CGB2, RMRP-CCDC 107, tRNA(Lys)-ALOXE3, RNU6-9-MED16: tRNA (Gly)-DPP9, RNU6-2-THEM259, and SNORD13-C8orf41,

wherein

a) one side of the bidirectional promoter is operably linked to at least one nucleotide sequence encoding a nuclease system guide RNA (gRNA), wherein the gRNA hybridizes with a target sequence of a DNA molecule or RNA molecule in a cell; and

b) the other side of the bidirectional promoter is operably linked to a nucleotide sequence encoding an RNA-directed nuclease,

wherein the gRNA targets and hybridizes with the target sequence and the nuclease cleaves one or both strands of the DNA molecule or RNA molecule.

2. The system of claim 1, wherein the system is CRISPR.

3. The system of claim 1, wherein the system is packaged into a single adeno-associated virus (AAV) particle.

4. The system of claim 1, wherein the system inactivates one or more gene products.

5. The system of claim 1, wherein the nuclease system excises at least one gene mutation.

6. The system of claim 1, further comprising a donor template.

7. The system of claim 2, wherein the nucleotide sequence encoding the RNA-directed nuclease is mutated so that the encoded RNA-directed nuclease lacks the ability to cleave one or both strands of the target polynucleotide.

8. The system of claim 1, further comprising at least one Kozak sequence.

9. The system of claim 1, further comprising an RNA sequence that mediates cap-independent initiation of translation.

10. The system of claim 1, wherein the promoter is a 7sk promoter.

11. The system of claim 10, wherein the 7sk promoter comprises a nucleotide sequence having at least 95% identity to the nucleotide sequence set forth in SEQ ID NOs: 3-8.

12. The system of claim 11, wherein the 7sk promoter comprises a nucleotide sequence set forth in SEQ ID NOs: 3-8.

13. A non-naturally occurring nuclease system comprising a vector comprising:

a) an RPPH1-PARP2 bidirectional promoter selected from *bos taurus, chlorocebus sabaeus, dasypus novemcinctus, oryctolagus cuniculus, pan troglodytes, papio anubis, tursiops truncatus,* and *vicugna pacos*, wherein the bidirectional promoter comprises:

a) control elements that provide for transcription in one direction of at least one nucleotide sequence encoding a nuclease system guide RNA (gRNA), wherein the gRNA hybridizes with a target sequence of a DNA molecule or RNA molecule in a cell; and

b) control elements that provide for transcription in the opposite direction of a nucleotide sequence encoding an RNA-directed nuclease,

wherein the gRNA targets and hybridizes with the target sequence and the nuclease cleaves one or both strands of the DNA molecule or RNA molecule.

14. The system of claim 13, wherein the system is CRISPR.

15. The system of claim 13, wherein the system is packaged into a single adeno-associated virus (AAV) particle.

16. The system of claim 13, wherein the system inactivates one or more gene products.

17. The system of claim 13, wherein the nuclease system excises at least one gene mutation.

18. The system of claim 13, further comprising a donor template.

19. The system of claim 13, wherein the nucleotide sequence encoding the RNA-directed nuclease is mutated so that the encoded RNA-directed nuclease lacks the ability to cleave one or both strands of the target polynucleotide.

20. The system of claim 13, further comprising at least one Kozak sequence.

21. The system of claim 13, further comprising an RNA sequence that mediates cap-independent initiation of translation.

22. The system of any preceding claim, wherein the RNA-directed nuclease is a Cas9 nickase derived from *S. aureus* (SaCas9 nickase).

23. Use of the non-naturally occurring nuclease system of claim 1 in the manufacture of a medicament for the treatment of a disease selected from the group consisting of retinal dystrophy, corneal dystrophy, microsatellite expansion diseases, and cancer in a subject in need thereof, wherein the disease results from disruption in a disease-associated gene.

24. The non-naturally occurring nuclease system of claim 1 for use in the treatment of a disease selected from the group consisting of retinal dystrophy, corneal dystrophy, microsatellite expansion diseases, and cancer in a subject in need thereof, wherein the disease results from disruption in a disease-associated gene.

25. Use of the non-naturally occurring nuclease system of claim 13 in the manufacture of a medicament for the treatment of a disease selected from the group consisting of retinal dystrophy, corneal dystrophy, microsatellite expansion diseases, and cancer in a subject in need thereof, wherein the disease results from disruption in a disease-associated gene.

26. The non-naturally occurring nuclease system of claim 13 for use in the treatment of a disease selected from the group consisting of retinal dystrophy, corneal dystrophy, microsatellite expansion diseases, and cancer in a subject in need thereof, wherein the disease results from disruption in a disease-associated gene.