DIRECTED EDITING OF CELLULAR RNA VIA NUCLEAR DELIVERY OF CRISPR/CAS9

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application claims priority under 35 U.S.C. § 119(e) to U.S. Provisional Application No. 62/504,497 filed May 10, 2017, the content of which is incorporated herein by reference in its entirety.

STATEMENT OF GOVERNMENT SUPPORT

[0002] This invention was made with government support under Grant Nos. HG004659 and NS075449 awarded by the National Institutes of Health. The government has certain rights in the invention.

BACKGROUND

[0003] Present strategies aimed to target and manipulate RNA in living cells mainly rely on the use of antisense oligonucleotides (ASO) or engineered RNA binding proteins (RBP). Although ASO therapies have shown great promise in eliminating pathogenic transcripts or modulating RBP binding, they are synthetic in construction and thus cannot be encoded within DNA. This complicates potential gene therapy strategies, which would rely on regular administration of ASOs throughout the lifetime of the patient. Furthermore, they are incapable of modulating the genetic sequence of RNA. Although RBPs such as the Pumilio and FBF homology family (PUF) of proteins can be designed to recognize target transcripts and fuse to RNA modifying effectors to allow for specific recognition and manipulation, platforms based on these types of constructs require extensive protein engineering for each target and may prove to be difficult and costly.

[0004] Current systems used to directly edit RNA rely either on non encodable components, such as chemical fusion of guide RNAs to an editase moiety (e.g., SNAP tag), or relatively low affinity tethering by fusion of encodable aptamer binding moieties (e.g., BoxB protein).

[0005] Current CRISPR/Cas RNA targeting systems typically use a single guide RNA and optionally an oligonucleotide of alternating 2' OMe RNA and DNA bases (PAMmer) to provide a simple and rapidly programmable system for targeting of specific RNA molecules in live cells. However, improvements and/or alternatives to these systems can help address issues relating to efficiency, specificity and/or off-target editing events. The present disclosure addresses these needs and provides related advantages.

SUMMARY OF THE DISCLOSURE

[0006] Accordingly, provided herein are fully encodable and highly specific CRISPR/Cas systems, compositions, and methods to achieve efficient and reversible manipulation and modulation of target RNA with simplicity, reliability and versatility.

[0007] In some examples, provided herein are recombinant expression systems for CRISPR/Cas-directed RNA editing of a target RNA comprising, consisting of, or consisting essentially of: (A) a nucleic acid sequence encoding a CRISPR/Cas RNA editing fusion protein comprising a nuclease-dead CRISPR associated endonuclease (dCas) fused to a catalytically active deaminase domain of Adenosine Deaminase acting on RNA (ADAR); and (B) a nucleic acid sequence encoding an extended single guide RNA (esgRNA) comprising: (i) a short extension sequence of homology to the target RNA comprising a mismatch for a target adenosine, and (ii) a dCas scaffold binding sequence. In one aspect, the present disclosure refers to a recombinant expression system for CRISPR/Cas-directed RNA editing of a target RNA comprising: (A) a nucleic acid sequence encoding a CRISPR/Cas RNA editing fusion protein comprising a nuclease-dead CRISPR associated endonuclease (dCas) fused to a catalytically active deaminase domain of Adenosine Deaminase acting on RNA 2 (ADAR2); and (B) a nucleic acid sequence encoding an extended single guide RNA (esgRNA) comprising: (i) a short extension sequence of homology to the target RNA comprising a mismatch for a target adenosine, and (ii) a dCas9 scaffold binding sequence. In some examples, said expression system expresses a dCas-ADAR nucleoprotein complex capable of CRISPR/Cas RNA-RNA base-specific Adenosine to Inosine (A – I) editing of the target sequence.

[0008] In some examples of the recombinant expression systems, the esgRNA further comprises (iii) a spacer sequence comprising a region of homology to the target RNA.

[0009] In some examples of the recombinant expression systems, (A) and (B) are comprised within the same vector or comprised within different vectors. In some examples of the recombinant expression systems, the vector is a viral vector. In some examples of the recombinant expression systems, the viral vector is an adeno-associated viral vector (AAV), lentiviral vector, or an adenoviral vector.

[0010] In some examples of the recombinant expression systems, the ADAR is selected from the group consisting of ADAR1, ADAR2, and ADAR3. In some examples, the catalytically active deaminase domain of ADAR is the catalytically active deaminase domain of ADAR2. In some examples of the recombinant expression systems, the catalytically active deaminase domain of ADAR2 is (1) a wildtype catalytically active deaminase domain of human ADAR2 or (2) a mutant human catalytically active deaminase domain of ADAR2 with increased catalytic activity compared to the wildtype human ADAR2. In some examples of the recombinant expression systems, the mutant human catalytically active deaminase domain of ADAR2 comprises a E488Q mutation.

[0011] In some examples of the recombinant expression systems, the dCas is nuclease-dead Cas9 (dCas9). In some examples of the recombinant expression systems, the dCas9 N-terminal domain is fused to the C-terminus of the catalytically active deaminase domain of ADAR. In some examples of the recombinant expression systems, the dCas is fused to the catalytically active deaminase domain of ADAR via a linker. In some examples of the recombinant expression systems, the linker is a semi-flexible XTEN peptide linker. In some examples, the linker is a GSGS linker.

[0012] In some examples of the recombinant expression systems, the short extension sequence of the esgRNA is a 3' extension sequence. In some examples of the recombinant expression systems, the short extension sequence of the esgRNA comprises a region of homology capable of near-perfect RNA-RNA base pairing with the target sequence. In some examples of the recombinant expression systems, the short extension sequence of the esgRNA further comprises a second mismatch for an adenosine within the target RNA. In some examples of the recombinant expression systems, the short extension sequence of the esgRNA further comprises a third mismatch for an adenosine within the target RNA and optionally a fourth mismatch for an adenosine within the target RNA. In some examples of the recombinant expression systems, the short extension sequence of the esgRNA is about 15 nucleotides to about 60 nucleotides in length.

[0013] In some examples of the recombinant expression systems, the esgRNA further comprises a marker sequence.

[0014] In some examples of the recombinant expression systems, the esgRNA further comprises a RNA polymerase III promoter sequence. In some examples of the recombinant expression systems, the RNA polymerase III promoter sequence is a U6 promoter sequence.

[0015] In some examples of the recombinant expression systems, the esgRNA comprises a linker sequence between the spacer sequence and the scaffold sequence.

[0016] In some examples of the recombinant expression systems, the sequences of the esgRNA (i), (ii), and (iii) are situated 3' to 5' in the esgRNA.

[0017] In some examples of the recombinant expression systems, the expression system further comprises a nucleic acid encoding a PAM sequence.

[0018] In some examples, provided herein are vectors comprising, consisting of, or consisting essentially of a nucleic acid encoding an extended single guide RNA (esgRNA) comprising (i) a short extension sequence of homology to a target RNA comprising a mismatch for a target adenosine, (ii) a dCas scaffold binding sequence, and (iii) a sequence complementary to the target sequence (spacer sequence), wherein (i), (ii) and (iii) are situated 3' to 5' in the esgRNA. In one example, the present disclosure refers to a vector comprising a nucleic acid encoding an extended single guide RNA (esgRNA) comprising (i) a short extension sequence of homology to a target RNA comprising a mismatch for a target adenosine, (ii) a dCas scaffold binding sequence, and (iii) a sequence complementary to the target sequence (spacer sequence), wherein (i), (ii) and (iii) are situated 3' to 5' in the esgRNA.

[0019] In some examples of the vectors, the vector is a viral vector. In some examples of the vectors, the viral vector is an adeno-associated viral vector (AAV), lentiviral vector, or an adenoviral vector. In some examples of the vectors, the vectors further comprise an expression control element.

[0020] In one example, the present disclosure refers to a viral particle comprising any one of the vectors disclosed herein. In some examples, provided herein are viral particles comprising a vector comprising, consisting of, or consisting essentially of a nucleic acid encoding an extended single guide RNA (esgRNA) comprising (i) a short extension sequence of homology to a target RNA comprising a mismatch for a target adenosine, (ii) a dCas scaffold binding sequence, and (iii) a sequence complementary to the target sequence (spacer sequence), wherein (i), (ii) and (iii) are situated 3' to 5' in the esgRNA. In some examples, provided herein are viral particles comprising one or more vectors comprising (A) a nucleic acid sequence encoding a CRISPR/Cas RNA editing fusion protein comprising a nuclease-dead CRISPR associated

endonuclease (dCas) fused to a catalytically active deaminase domain of Adenosine Deaminase acting on RNA (ADAR); and (B) a nucleic acid sequence encoding an extended single guide RNA (esgRNA) comprising: (i) a short extension sequence of homology to the target RNA comprising a mismatch for a target adenosine, and (ii) a dCas scaffold binding sequence.

[0021] In one aspect, the present disclosure refers to a cell comprising any one of the recombinant expression systems disclosed herein. In some examples, provided herein are cells comprising recombinant expression systems, viral particles, and/or vectors comprising, consisting of, or consisting essentially of a nucleic acid encoding an extended single guide RNA (esgRNA) comprising (i) a short extension sequence of homology to a target RNA comprising a mismatch for a target adenosine, (ii) a dCas scaffold binding sequence, and (iii) a sequence complementary to the target sequence (spacer sequence), wherein (i), (ii) and (iii) are situated 3' to 5' in the esgRNA. In some examples, provided herein are cells comprising one or more viral particles, recombinant expression systems, and/or vectors comprising (A) a nucleic acid sequence encoding a CRISPR/Cas RNA editing fusion protein comprising a nuclease-dead CRISPR associated endonuclease (dCas) fused to a catalytically active deaminase domain of Adenosine Deaminase acting on RNA (ADAR); and (B) a nucleic acid sequence encoding an extended single guide RNA (esgRNA) comprising: (i) a short extension sequence of homology to the target RNA comprising a mismatch for a target adenosine, and (ii) a dCas scaffold binding sequence.

[0021a] In one example, the present disclosure refers to an esgRNA comprising: (i) a short extension sequence of homology to a target RNA comprising a mismatch for a target adenosine, and (ii) a dCas scaffold binding sequence. In another example, the present disclosure refers to a CRISPR/Cas RNA editing fusion protein comprising a nuclease-dead CRISPR associated endonuclease (dCas) fused to a catalytically active deaminase domain of ADAR.

[0022] In one aspect, the present disclosure refers to a method of selective RNA editing comprising administering any one of the recombinant expression systems as disclosed herein to a cell, wherein the method is *in vitro*. Also provided herein are methods of selective RNA editing comprising, consisting of, or consisting essentially of administering any one of the recombinant expression systems, viral particles, and/or vectors comprising, consisting of, or consisting essentially of a nucleic acid encoding an extended single guide RNA (esgRNA) comprising (i) a short extension sequence of homology to a target RNA comprising a mismatch for a target adenosine, (ii) a dCas scaffold binding sequence, and (iii) a sequence complementary to the target sequence (spacer sequence), wherein (i), (ii) and (iii) are situated 3' to 5' in the esgRNA to a cell. In some examples, the methods further comprise administering an antisense synthetic oligonucleotide compound comprising alternating 2'OMe RNA and DNA bases (PAMmer). In some examples, the method is *in vitro* or *in vivo*. In some examples, provided herein are methods of selective RNA editing comprising, consisting of, or consisting

essentially of administering any one of the recombinant expression systems, viral particles, and/or vectors comprising, consisting of, or consisting essentially of (A) a nucleic acid sequence encoding a CRISPR/Cas RNA editing fusion protein comprising a nuclease-dead CRISPR associated endonuclease (dCas) fused to a catalytically active deaminase domain of Adenosine Deaminase acting on RNA (ADAR); and (B) a nucleic acid sequence encoding an extended single guide RNA (esgRNA) comprising: (i) a short extension sequence of homology to the target RNA comprising a mismatch for a target adenosine, and (ii) a dCas scaffold binding sequence.

[0023] In one aspect, the present disclosure refers to a method of characterizing the effects of directed cellular RNA editing on processing and dynamics comprising administering any one of the recombinant expression systems as disclosed herein to a sample and determining its effects. Also provided herein are methods of characterizing the effects of directed cellular RNA editing on processing and dynamics comprising administering any one of the recombinant expression systems, viral particles, and/or vectors comprising, consisting of, or consisting essentially of a nucleic acid encoding an extended single guide RNA (esgRNA) comprising (i) a short extension sequence of homology to a target RNA comprising a mismatch for a target adenosine, (ii) a dCas scaffold binding sequence, and (iii) a sequence complementary to the target sequence (spacer sequence), wherein (i), (ii) and (iii) are situated 3' to 5' in the esgRNA to a sample and determining its effects. In some examples, the sample is derived from a subject. In some examples, the method is *in vitro* or *in vivo*. In some examples, provided herein are methods of characterizing the effects of directed cellular RNA editing on processing and dynamics comprising administering any one of the recombinant expression systems, viral particles, and/or vectors comprising, consisting of, or consisting essentially of (A) a nucleic acid sequence encoding a CRISPR/Cas RNA editing fusion protein comprising a nuclease-dead CRISPR associated endonuclease (dCas) fused to a catalytically active deaminase domain of Adenosine Deaminase acting on RNA (ADAR); and (B) a nucleic acid sequence encoding an extended single guide RNA (esgRNA) comprising: (i) a short extension sequence of homology to the target RNA comprising a mismatch for a target adenosine, and (ii) a dCas scaffold binding sequence to a sample and determining its effects.

[0024] In one aspect, the present disclosure refers to use of any one of the recombinant expression systems as disclosed herein, or the cell as disclosed herein, in the manufacture of a medicament for treating a disease selected from the group consisting of Hurler's syndrome, Cystic fibrosis, muscular dystrophy, spinal cord injury, hearing loss (through noise overexposure or ototoxicity), multiple sclerosis, Alzheimer's disease, amyotrophic lateral sclerosis (ALS), Parkinson's disease, over-rapid benzodiazepine withdrawal, and Huntington's disease in a subject, wherein the medicament is to be administered to a subject or a sample isolated from a subject. In other examples, provided herein are methods of treating a disease or condition in a subject comprising administering any one of the recombinant expression systems, viral particles, and/or vectors comprising, consisting of, or consisting essentially of a nucleic acid

encoding an extended single guide RNA (esgRNA) comprising (i) a short extension sequence of homology to a target RNA comprising a mismatch for a target adenosine, (ii) a dCas scaffold binding sequence, and (iii) a sequence complementary to the target sequence (spacer sequence), wherein (i), (ii) and (iii) are situated 3' to 5' in the esgRNA to a subject or a sample isolated from a subject. In some examples, provided herein are methods of treating a disease or condition in a subject comprising administering any one of the recombinant expression systems, viral particles, and/or vectors comprising, consisting of, or consisting essentially of (A) a nucleic acid sequence encoding a CRISPR/Cas RNA editing fusion protein comprising a nuclease-dead CRISPR associated endonuclease (dCas) fused to a catalytically active deaminase domain of Adenosine Deaminase acting on RNA (ADAR); and (B) a nucleic acid sequence encoding an extended single guide RNA (esgRNA) comprising: (i) a short extension sequence of homology to the target RNA comprising a mismatch for a target adenosine, and (ii) a dCas scaffold binding sequence to a subject or a sample isolated from a subject.

[0025] In some examples, the methods further correcting a G to A mutation in a target RNA. In some examples, the disease is selected from the group of Hurler's syndrome, Cystic fibrosis, Duchenne muscular dystrophy, spinal cord injury, stroke, traumatic brain injury, hearing loss (through noise overexposure or ototoxicity), multiple sclerosis, Alzheimer's disease, amyotrophic lateral sclerosis (ALS), Parkinson's disease, alcoholism, alcohol withdrawal, over-rapid benzodiazepine withdrawal, and Huntington's disease.

[0026] In one aspect, the present disclosure refers to a kit comprising: (A) one or more of: (i) a recombinant expression system as disclosed herein; (ii) a viral particle comprising the recombinant expression system as disclosed herein; and/or (iii) a cell comprising the recombinant expression system as disclosed herein; and (B) instructions for use. In one example, the present disclosure refers to a kit comprising: (A) one or more of: (i) a recombinant expression system as disclosed herein; (ii) a vector comprising a nucleic acid encoding an esgRNA comprising (a) a short extension sequence of homology to a target RNA comprising a mismatch for a target adenosine, and (b) a dCas scaffold binding sequence; (iii) a vector comprising a nucleic acid encoding a dCas fused to a catalytically active deaminase domain of ADAR; (iv) a viral particle comprising the recombinant expression system as disclosed herein; (v) a cell comprising the recombinant expression system as disclosed herein; (vi) an esgRNA comprising (a) a short extension sequence of homology to a target RNA comprising a mismatch for a target adenosine, and (b) a dCas scaffold binding sequence; and/or (vii) a CRISPR/Cas RNA editing fusion protein comprising a dCas fused to a catalytically active deaminase domain of ADAR; and (B) instructions for use. In other examples, provided herein are kits comprising, consisting of, or consisting of one or more of: recombinant expression systems, viral particles, and/or vectors comprising, consisting of, or consisting essentially of (A) a nucleic acid sequence encoding a CRISPR/Cas RNA editing fusion protein comprising a nuclease-dead CRISPR associated endonuclease (dCas) fused to a catalytically active deaminase domain of Adenosine Deaminase acting on RNA (ADAR); and (B) a nucleic acid sequence encoding an extended single guide RNA (esgRNA) comprising: (i) a short extension sequence of homology to the target RNA comprising a mismatch for a target adenosine, and (ii) a dCas scaffold binding sequence and instructions for use. In some examples, the instructions are for use according to any one of the methods described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0027] **FIGs. 1A-1D** illustrate, without limitation, embodiments of the recombinant expression system and data relating thereto. **FIG. 1A** shows (i) a conceptual concept of CREDIT in living cells for the editing of a variety of RNAs that can cause various diseases, such as cancer and neurodegeneration and (ii) that the binding of the dCas9-deaminase fusion to guide RNA directs the hybridization of guide-extension around target adenosines generating double-stranded RNA (dsRNA) A-I base-specific editing targets. In particular, **FIG. 1B** shows a CREDIT recombinant expression system comprised of the *Streptococcus pyogenes* Cas9 protein fused by an XTEN linker to the deaminase domain (DD) of human ADARB1 (ADAR2), and a single guide RNA (sgRNA) with a 3' short RNA extension (esgRNA). The fluorescent imaging data of **FIG. 1C** shows that the recombinant expression system of Figure 1B requires targeted dual guide RNA with 3' extension directing deamination and allows reversal of premature termination codon (PTC) mediated silencing of expression from eGFP reporter transcripts. **FIG. 1D** shows FACS quantification of recombinant expression systems utilizing wild-type and hyper-active deaminase fusions to RCas9 directed by targeting and non-targeting guides.

[0028] **FIG. 2** illustrates, without limitation, an exemplary recombinant expression system as an AAV-based vector system. The AAV system comprises vectors carrying the nucleic acid sequence encoding the ADAR Deaminase domain/ Cas endonuclease fusion protein and the extended single guide RNA (esgRNA) to be packaged as AAV virions.

[0029] **FIG. 3** illustrates a map of pcDNA3.1(1)_ADAR2_XTEN_dCas9 (SEQ ID NO: 27). The CMV enhancer is located at postion 235 to 614 (380bp in length) and drives constitutive expression of recombinant protein in mammalian cells. The CMV promoter is located at postion 615 to 818 (204 bp in length) and drives constitutive expression of recombinant protein in mammalian cells. The ADARB1 Catalytic Domain is located at position 961 to 2100 (1140 bp in length) and encodes a catalytically-active deaminating domain of human ADAR2 (ADARB1). XTEN is located at position 2101 to 2148 (48bp in length) and encodes a peptide linker connecting recombinant protein domains. dCas9 is located at postion 2149 to 6252 (4104 bp in length) and encodes a catalytically-inactive

(D10A and H841A) CRISPR-Cas9 protein from *Streptococcus pyogenes*. HA is located at postion 6256 to 6282 (27 bp in length) and encodes human influenza hemagglutinin (HA) epitope tag. 2X SV40 NLS is located at postion 6301 to 6348 (48 bp in length) and encodes a Nuclear localization signal (NLS) derived from Simian Virus 40 (SV40) large T-antigen. bGH poly(A) signal is located at postion 6426 to 6650 (225 bp in length) and encodes a bovine growth hormone (bGH) polyadenylation signal.

[0030] **FIG. 4** illustrates a map of pcDNA3.1(1)_ADAR2_XTEN_control (SEQ ID NO: 28). A CMV enhancer is located at position 235 to 614 (380 bp in length) and drives constitutive expression of recombinant protein in mammalian cells. A CMV promoter is located at position 615 to 818 (204 bp in length) and drives constitutive expression of recombinant protein in mammalian cells. An ADARB1 Catalytic Domain is located at position 961 to 2100 (1140 bp in length) and encodes a catalytically-active deaminating domain of human ADAR2 (ADARB1). XTEN is located at position 2101 to 2148 (48 bp) and encodes a peptide linker connecting recombinant protein domains. HA is located at position 2152 to 2178 (27 bp) and encodes human influenza hemagglutinin (HA) epitope tag 2X SV40 NLS is located at position 2197 to 2244 (48bp) nuclear localization signal (NLS) derived from Simian Virus 40 (SV40) large T-antigen. bGH poly(A) signal is located at position 2322 to 2546 (225 bp) and encodes bovine growth hormone (bGH) polyadenylation signal.

[0031] **FIG. 5** illustrates a map of pcDNA3.1_ADAR2(E488Q)_XTEN_dCas9 (SEQ ID NO: 29). A CMV enhancer is located at position 235 to 614 (380 bp) and drives constitutive expression of recombinant protein in mammalian cells. A CMV promoter is located at position 615 to 818 (204 bp) and drives constitutive expression of recombinant protein in mammalian cells. ADARB1(E488Q) Catalytic Domain is located at position 961 to 2100 (1140 bp) and encodes a catalytically-active deaminating domain of human ADAR2 (ADARB1) with hyperactive point mutation (E488Q). XTEN is located at position 2101 to 2148 (48 bp) and encodes a peptide linker connecting recombinant protein domains. dCas9 is located at position 2149 to 6252 (4104 bp) and encodes a catalytically-inactive (D10A and H841A) CRISPR-Cas9 protein from *Streptococcus pyogenes*. HA is located at position 6256 to 6282 (27 bp) and encodes human influenza hemagglutinin (HA) epitope tag. 2X SV40

NLS is located at position 6301 to 6348 (48 bp) and encodes a nuclear localization signal (NLS) derived from Simian Virus 40 (SV40) large T-antigen bGH. poly(A) signal is located at position 6426 to 6650 (225 bp) and encodes bovine growth hormone (bGH) polyadenylation signal.

[0032] **FIG. 6** illustrates a map of pcDNA3.1_ADAR2(E488Q)_XTEN_control (SEQ ID NO: 30). A CMV enhancer is located at position 235 to 614 (380bp) and drives constitutive expression of recombinant protein in mammalian cells. A CMV promoter is located at position 615 to 818 (204 bp) and drives constitutive expression of recombinant protein in mammalian cells. ADARB1(E488Q) Catalytic Domain is located at position 961 to 2100 (1140 bp) and encodes a catalytically-active deaminating domain of human ADAR2 (ADARB1) with hyperactive point mutation (E488Q). XTEN is located at position 2101 to 2148 (48 bp) and encodes a peptide linker connecting recombinant protein domains. HA is located at position 2152 to 2178 (27 bp) and encodes a human influenza hemagglutinin (HA) epitope tag. 2X SV40 NLS is located at position 2197 to 2244 (48 bp) and encodes a nuclear localization signal (NLS) derived from Simian Virus 40 (SV40) large T-antigen. bGH poly(A) signal is located at position 2322 to 2546 (225 bp) and encodes bovine growth hormone (bGH) polyadenylation signal.

[0033] **FIG. 7** illustrates a map of 50bp_GFP_mCherry_extension (SEQ ID NO: 31). A U6 promoter is located at position 4555 to 4817 (263 bp) and is a Pol III promoter driving expression of sgRNA in mammalian cells. An EGFP targeting spacer is located at position 4818 to 4838 (21 bp) and encodes a spacer sequence of sgRNA that targets complementary EGFP reporter mRNA. An sgRNA scaffold is located at position 4839 to 4924 (86 bp) and encodes an sgRNA scaffold for *Streptococcus pyogenes* CRISPR-Cas9 system with (F+E) modification (Chen et al. 2014). Linker is located at position 4925 to 4930 (6 bp) encoding a linker sequence bridging the sgRNA scaffold with the extension sequence. And EGFP extension is located at position 4931 to 4951 (21 bp) encoding an RNA extension sequence that base pairs with target site and forces A-to-I editing using A-C mismatch. A sgRNA scaffold termination site is located at position 1 to 7 (7 bp) comprising a Poly(T) sequence that terminates Pol III RNA synthesis. An Ef1a promoter is located at position 21 to 566 (546 bp) which is a constitutive promoter driving protein expression in mammalian cells. mCherry

is located at position 572 to 1282 (711 bp) encoding a monomeric derivative of DsRed fluorescent protein. A bGH poly(A) signal is located at position 1330 to 1554 (225 bp) encoding a bovine growth hormone (bGH) polyadenylation signal.

[0034] FIG. 8 illustrates a map of spacerless_GFP_mCherry_extension (SEQ ID NO: 32). A U6 promoter is located at position 757 to 1019 (263 bp) and is a Pol III promoter driving expression of sgRNA in mammalian cells. An sgRNA scaffold is located at position 1020 to 1105 (86 bp) encoding an sgRNA scaffold for *Streptococcus pyogenes* CRISPR-Cas9 system with (F+E) modification (Chen et al. 2014). A Linker is located at position 1106 to 1111 (6 bp) comprising a linker sequence bridging the sgRNA scaffold with the extension sequence. An EGFP extension is located at position 1112 to 1132 (21 bp) encoding an RNA extension sequence that base pairs with target site and forces A-to-I editing using A-C mismatch. An sgRNA scaffold termination is located at position 1133 to 1139 (7 bp) comprising a poly(T) sequence that terminates Pol III RNA synthesis. An Ef1a promoter is located at position 1153 to 1698 (546 bp) and is a constitutive promoter driving protein expression in mammalian cells. mCherry is located at position 1704 to 2414 (711 bp) encoding a monomeric derivative of DsRed fluorescent protein. A bGH poly(A) signal is located at position 2462 to 2686 (225 bp) encoding bovine growth hormone (bGH) polyadenylation signal.

[0035] FIG. 9 illustrates a map of GFP_no_spacer_revcomp_mCherry_gibson (SEQ ID NO: 33). A U6 promoter is located at position 4555 to 4817 (263 bp) and is a Pol III promoter driving expression of sgRNA in mammalian cells. An sgRNA scaffold is located at position 4818 to 4903 (86 bp) and encodes a sgRNA scaffold for *Streptococcus pyogenes* CRISPR-Cas9 system with (F+E) modification (Chen et al. 2014). A linker is located at position 4904 to 4909 (6 bp) encoding a linker sequence bridging the sgRNA scaffold with the extension sequence. An EGFP revcomp extension is located at position 4910 to 4930 (21 bp) encoding an RNA reverse complement extension sequence that matches the sequence of the EGFP mRNA target site. An sgRNA scaffold termination site is located at position 1 to 7 (7 bp) comprising a poly(T) sequence that terminates Pol III RNA synthesis. An Ef1a promoter is located at position 21 to 566 (546 bp) and is a constitutive promoter driving protein expression in mammalian cells. mCherry is located at position 572 to 1282 (711 bp) encoding a monomeric derivative of DsRed fluorescent protein. A bGH poly(A) signal is

located at position 1330 to 1554 (225 bp) encoding a bovine growth hormone (bGH) polyadenylation signal.

[0036] FIG. 10 illustrates a map of pBluescript II SK+ U6-lambda2-sgRNA(F+E) (SEQ ID NO: 34). A U6 promoter is located at position 757 to 1019 (263 bp) and is a Pol III promoter driving expression of sgRNA in mammalian cells. A lambda2 guideRNA is located at position 1020 to 1039 (20 bp) encoding a non-targeting sgRNA sequence targeting lambda phage 2. An sgRNA scaffold is located at position 1041 to 1132 (92 bp) encoding a sgRNA scaffold for *Streptococcus pyogenes* CRISPR-Cas9 system with (F+E) modification (Chen et al. 2014).

[0037] FIG. 11 illustrates a map of EGFP_spacerless_SaCas9_sgRNA (SEQ ID NO: 47). A U6 promoter is located at position 4555 to 4817 (263 bp) and is a Pol III promoter driving expression of sgRNA in mammalian cells. An *Sa* sgRNA scaffold is located at position 4819 to 4894 (76 bp) encoding an sgRNA scaffold for *Staphylococcus aureus* CRISPR-Cas9 system with A-U base flip (Chen et al. 2016). A linker is located at position 4895 to 4900 (6 bp) encoding a linker sequence bridging the sgRNA scaffold with the extension sequence. An EGFP extension is located at position 4901 to 4921 (21 bp) encoding an RNA extension sequence that base pairs with target site and forces A-to-I editing using A-C mismatch. An sgRNA scaffold termination site is located at position 1 to 7 (7 bp) comprising a poly(T) sequence that terminates pol III RNA synthesis. An Ef1a promoter is located at position 21 to 566 (546 bp) which is a constitutive promoter driving protein expression in mammalian cells. mCherry is located at position 572 to 1282 (711 bp) encoding a monomeric derivative of DsRed fluorescent protein. A bGH poly(A) signal is located at position 1330 to 1554 (225 bp) encoding bovine growth hormone (bGH) polyadenylation signal.

[0038] FIG. 12 illustrates a map of ADAR2_E488Q_dSaCas9_pCDNA3_1 (SEQ ID NO: 48). A CMV enhancer is located at position 235 to 614 (380 bp) and drives constitutive expression of recombinant protein in mammalian cells. A CMV promoter is located at position 615 to 818 (204 bp) and drives constitutive expression of recombinant protein in mammalian cells. ADARB1 Catalytic Domain is located at position 961 to 2100 (1140 bp) and encodes a catalytically-active deaminating domain of human ADAR2 (ADARB1). A GS linker is located at position 2101 to 2112 (12 bp) and encodes a Glycine-Serine peptide linker

to bridge protein domains. A dSaCas9 is located at position 2113 to 5268 (3156 bp) encoding a catalytically-inactive (with point mutations D10A and N580A) CRISPR-Cas9 protein from *Staphylococcus aureus*. HA is located at position 5272 to 5298 (27 bp) encoding human influenza hemagglutinin (HA) epitope tag. A 2X SV40 NLS is located at position 5317 to 5364 (48 bp) nuclear localization signal (NLS) derived from Simian Virus 40 (SV40) large T-antigen. A bGH poly(A) signal is located at position 5442 to 5666 (225 bp) encoding a bovine growth hormone (bGH) polyadenylation signal.

[0039] **FIGs. 13A-13B** illustrate a comparison between a recombinant expression system comprising a nuclease dead Cas9 derived from S. pyogenes (dSpCas9) and a nuclease dead Cas9 derived from S. aureus (dSaCas9). dSaCas9 is significantly smaller than dSpCas9, which provides efficiency in viral packaging. **FIG. 13A** shows an illustration of an ADAR2(E488Q)-dSpCas9 fusion construct with an XTEN linker (Sp-CREDITv1) and an illustration of an ADAR2(E488Q)-dSaCas9 fusion construct with an GSGS linker (Sa-CREDITv1). **FIG. 1B** shows the results of an experiment wherein the efficiency of Sp-CREDITv1 is compared to the efficiency of Sa-CREDITv1. This data shows successful editing of the GFP reporter by both CREDIT systems, with Sa-CREDITv1 exhibiting the highest frequency of edited cells.

DETAILED DESCRIPTION

[0040] Embodiments according to the present disclosure will be described more fully hereinafter. Aspects of the disclosure may, however, be embodied in 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 be thorough and complete, and will fully convey the scope of the invention to those skilled in the art. The terminology used in the description herein is for the purpose of describing particular embodiments only and is not intended to be limiting.

[0041] Unless otherwise defined, all terms (including technical and scientific terms) used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. It will be further understood that terms, such as those defined in commonly used dictionaries, should be interpreted as having a meaning that is consistent

with their meaning in the context of the present application and relevant art and should not be interpreted in an idealized or overly formal sense unless expressly so defined herein. While not explicitly defined below, such terms should be interpreted according to their common meaning.

[0042] The terminology used in the description herein is for the purpose of describing particular embodiments only and is not intended to be limiting of the invention. All publications, patent applications, patents and other references mentioned herein are incorporated by reference in their entirety.

[0043] The practice of the present technology will employ, unless otherwise indicated, conventional techniques of tissue culture, immunology, molecular biology, microbiology, cell biology, and recombinant DNA, which are within the skill of the art.

[0044] Unless the context indicates otherwise, it is specifically intended that the various features of the invention described herein can be used in any combination. Moreover, the disclosure also contemplates that in some embodiments, any feature or combination of features set forth herein can be excluded or omitted. To illustrate, if the specification states that a complex comprises components A, B and C, it is specifically intended that any of A, B or C, or a combination thereof, can be omitted and disclaimed singularly or in any combination.

[0045] Unless explicitly indicated otherwise, all specified embodiments, features, and terms intend to include both the recited embodiment, feature, or term and biological equivalents thereof.

[0046] All numerical designations, e.g., pH, temperature, time, concentration, and molecular weight, including ranges, are approximations which are varied (+) or (-) by increments of 1.0 or 0.1, as appropriate, or alternatively by a variation of +/- 15 %, or alternatively 10%, or alternatively 5%, or alternatively 2%. It is to be understood, although not always explicitly stated, that all numerical designations are preceded by the term “about”. It also is to be understood, although not always explicitly stated, that the reagents described herein are merely exemplary and that equivalents of such are known in the art.

Definitions

[0047] As used in the description of the invention and the appended claims, the singular forms "a," "an" and "the" are intended to include the plural forms as well, unless the context clearly indicates otherwise.

[0048] The term "about," as used herein when referring to a measurable value such as an amount or concentration and the like, is meant to encompass variations of 20%, 10%, 5%, 1%, 0.5%, or even 0.1 % of the specified amount.

[0049] The terms or "acceptable," "effective," or "sufficient" when used to describe the selection of any components, ranges, dose forms, etc. disclosed herein intend that said component, range, dose form, etc. is suitable for the disclosed purpose.

[0050] "Polynucleotide" or "nucleotide," as used interchangeably herein, refer to polymers of nucleotides of any length, and include DNA and RNA. A polynucleotide or nucleotide sequence could be either double-stranded or single-stranded. When a polynucleotide or nucleotide sequence is single stranded, it could refer to either of the two complementary strands. The nucleotides can be deoxyribonucleotides, ribonucleotides, modified nucleotides or bases, and/or their analogs, or any substrate that can be incorporated into a polymer by DNA or RNA polymerase. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and their analogs. If present, modification 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. Other types of modifications include, for example, "caps", substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as, for example, those with uncharged linkages (such as methyl phosphonates, phosphotriesters, phosphoamidates, cabamates, etc.) and with charged linkages (such as phosphorothioates, phosphorodithioates, etc.), those containing pendant moieties, such as, for example, proteins (such as nucleases, toxins, antibodies, signal peptides, ply-L-lysine, etc.), those with intercalators (such as acridine, psoralen, etc.), those containing chelators (such as metals, radioactive metals, boron, oxidative metals, etc.), those containing alkylators, those with modified linkages (such as alpha anomeric nucleic acids, etc.), as well as unmodified forms

of the polynucleotide(s). Further, any of the hydroxyl groups ordinarily present in the sugars may be replaced, for example, by phosphonate groups, phosphate groups, protected by standard protecting groups, or activated to prepare additional linkages to additional nucleotides, or may be conjugated to solid supports. The 5′ and 3′ terminal OH can be phosphorylated or substituted with amines or organic capping groups moieties of from 1 to 20 carbon atoms. Other hydroxyls may also be derivatized to standard protecting groups. Polynucleotides can also contain analogous forms of ribose or deoxyribose sugars that are generally known in the art, including, for example, 2′-O-methyl-2′-O-allyl, 2′-fluoro- or 2′-azido-ribose, carbocyclic sugar analogs, α-anomeric sugars, epimeric sugars such as arabinose, xyloses or lyxoses, pyranose sugars, furanose sugars, sedoheptuloses, acyclic analogs and abasic nucleoside analogs such as methyl riboside. One or more phosphodiester linkages may be replaced by alternative linking groups. These alternative linking groups include, but are not limited to, embodiments wherein phosphate is replaced by P(O)S("thioate"), P(S)S ("dithioate"), "(O)NR 2 ("amidate"), P(O)R, P(O)OR′, CO or CH 2 ("formacetal"), in which each R or R′ is independently H or substituted or unsubstituted alkyl (1-20 C) optionally containing an ether (—O—) linkage, aryl, alkenyl, cycloalkyl, cycloalkenyl or araldyl. Not all linkages in a polynucleotide need be identical. The preceding description applies to all polynucleotides referred to herein, including RNA and DNA.

[0051] "Oligonucleotide," as used herein, generally refers to short, generally single stranded, generally synthetic polynucleotides that are generally, but not necessarily, less than about 200 nucleotides in length. The terms "oligonucleotide" and "polynucleotide" are not mutually exclusive. The description above for polynucleotides is equally and fully applicable to oligonucleotides.

[0052] "Nucleic acids", "nucleic acid molecules," or "nucleic acid sequences" are used interchangeably herein to refer to polynucleotides and/or oligonucleotides. In some embodiments, nucleic acid is used interchangeably with polynucleotide and/or oligonucleotide.

[0053] As used herein, "substantially complementary or substantially matched" means that two nucleic acid sequences have at least 90% sequence identity. Preferably, the two nucleic acid sequences have at least 95%, 96%, 97%, 98%, 99% or 100% of sequence identity.

Alternatively, "substantially complementary or substantially matched" means that two nucleic acid sequences can hybridize under high stringency condition(s).

[0054] As used herein, "improve" means a change of at least about 1%, 2%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 35%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, 125%, 150%, 175%, 200%, 225%, 250%, 275%, 300%, 350%, 400%, 450%, 500%, 600%, 700%, 800%, 900%, 1000% or more or any value between any of the listed values. Alternatively, "improve" could mean a change of at least about 1-fold, 1.5-fold, 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, 10-fold, 15-fold, 20-fold, 30-fold, 40-fold, 50-fold, 60-fold, 70-fold, 80-fold, 90-fold, 100-fold, 500-fold, 1000-fold, 2000-fold or more or any value between any of the listed values.

[0055] As used herein, "nuclease null" or "nuclease dead" may refer to a polypeptide with reduced nuclease activity, reduced endo- or exo-DNAse activity or RNAse activity, reduced nickase activity, or reduced ability to cleave DNA and/or RNA. Non-limiting examples of Cas-associated endonucleases that are nuclease dead include endonucleases with mutations that render the RuvC and/or HNH nuclease domains inactive. For example, *S. pyogenes* Cas9 can be rendered inactive by point mutations D10A and H840A, resulting in a nuclease dead Cas9 molecule that cannot cleave target DNA or RNA. The dCas9 molecule retains the ability to bind to target RNA based on the gRNA targeting sequence.

[0056] As used herein, "reduced nuclease activity" means a decline in nuclease, nickase, DNAse, or RNAse activity of at least about 1%, 2%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 35%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100% or more or any value between any of the listed values. Alternatively, "reduced nuclease activity" may refer to a decline of at least about 1-fold, 1.5-fold, 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, 10-fold, 15-fold, 20-fold, 30-fold, 40-fold, 50-fold, 60-fold, 70-fold, 80-fold, 90-fold, 100-fold, 500-fold, 1000-fold, 2000-fold or more or any value between any of the listed values.

[0057] As used herein, "increased catalytic activity" means an increase in catalytic activity of e.g. deaminase activity of at least about 1%, 2%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 35%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100% or more or any value between any of the listed values as compared to the corresponding wild type catalytic

activity (e.g., wild type deaminase activity). Alternatively, "increased catalytic activity" may refer to an increase of at least about 1-fold, 1.5-fold, 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, 10-fold, 15-fold, 20-fold, 30-fold, 40-fold, 50-fold, 60-fold, 70-fold, 80-fold, 90-fold, 100-fold, 500-fold, 1000-fold, 2000-fold or more or any value between any of the listed values as compared to the corresponding wild type catalytic activity (e.g., wild type deaminase activity).

[0058] As used herein, the term "ADAR" refers to a double-stranded RNA specific adenosine deaminase which catalyzes the hydrolytic deamination of adenosine to inosine in double-stranded RNA (dsRNA), referred to as A to I editing and also known as Adenosine Deaminase Acting on RNA. Non-limiting exemplary sequences of this protein and annotation of its domains is found under UniProt reference number P55265 (human) and Q99MU3 (mouse).

[0059] The term "adeno-associated virus" or "AAV" as used herein refers to a member of the class of viruses associated with this name and belonging to the genus dependoparvovirus, family Parvoviridae. Multiple serotypes of this virus are known to be suitable for gene delivery; all known serotypes can infect cells from various tissue types. At least 11, sequentially numbered, are disclosed in the prior art. Non-limiting exemplary serotypes useful in the methods disclosed herein include any of the 11 serotypes, e.g., AAV2 and AAV8.

[0060] Also as used herein, "and/or" refers to and encompasses any and all possible combinations of one or more of the associated listed items, as well as the lack of combinations when interpreted in the alternative ("or").

[0061] The term "aptamer" as used herein refers to single stranded DNA or RNA molecules that can bind to one or more selected targets with high affinity and specificity. Non-limiting exemplary targets include but are not limited to proteins or peptides.

[0062] The term "Cas-associated" refers to a CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) associated endonuclease. "Cas9" is a Cas-associated endonuclease referred to by this name (UniProtKB G3ECR1 (CAS9_STRTR)). DeadCas-9 or "dCas9" is a Cas9 endonuclease which lacks or substantially lacks endonuclease and/or

cleavage activity. A non-limiting example of dCas9 is the dCas9 encoded in AddGene plasmid .#74710, which is commercially available through the AddGene database.

[0063] The term "cell" as used herein may refer to either a prokaryotic or eukaryotic cell, optionally obtained from a subject or a commercially available source.

[0064] The term "gRNA" or "guide RNA" as used herein refers to the guide RNA sequences used to target specific genes for correction employing the CRISPR technique. Techniques of designing gRNAs and donor therapeutic polynucleotides for target specificity are well known in the art. For example, Doench, J., et al. Nature biotechnology 2014; 32(12):1262-7 and Graham, D., et al. Genome Biol. 2015; 16: 260, incorporated by reference herein.

[0065] As used herein, the term "CRISPR" refers to a technique of sequence specific genetic manipulation relying on the clustered regularly interspaced short palindromic repeats pathway, which unlike RNA interference regulates gene expression at a transcriptional level. The term "gRNA" or "guide RNA" as used herein refers to the guide RNA sequences used to target specific genes for correction employing the CRISPR technique. Techniques of designing gRNAs and donor therapeutic polynucleotides for target specificity are well known in the art. For example, Doench, J., et al. Nature biotechnology 2014; 32(12):1262-7 and Graham, D., et al. Genome Biol. 2015; 16: 260. "Single guide RNA" or "sgRNA" is a specific type of gRNA that combines tracrRNA (transactivating RNA), which binds to Cas9 to activate the complex to create the necessary strand breaks, and crRNA (CRISPR RNA), comprising complimentary nucleotides to the tracrRNA, into a single RNA construct. As described herein, an "extended single guide RNA" or "esgRNA" is a specific type of sgRNA that includes an extension sequence of homology to the target RNA comprising a mismatch for a target adenosine of the target RNA to be edited in a manner such that a A-C mismatch is formed with a target transcript generating a 'pseudo-dsRNA' substrate to be edited at the bulged adenosine residue.

[0066] As used herein, the term "comprising" is intended to mean that the compositions and methods include the recited elements, but do not exclude others. As used herein, the transitional phrase "consisting essentially of' (and grammatical variants) is to be interpreted as encompassing the recited materials or steps "and those that do not materially affect the

basic and novel characteristic(s)" of the recited embodiment. *See, In re Herz,* 537 F.2d 549, 551-52, 190 U.S.P.Q. 461, 463 (CCPA 1976) (emphasis in the original); *see also* MPEP § 2111.03. Thus, the term "consisting essentially of' as used herein should not be interpreted as equivalent to "comprising." "Consisting of" shall mean excluding more than trace elements of other ingredients and substantial method steps for administering the compositions disclosed herein. Aspects defined by each of these transition terms are within the scope of the present disclosure.

[0067] The term "encode" as it is applied to nucleic acid sequences refers to a polynucleotide which is said to "encode" a polypeptide if, in its native state or when manipulated by methods well known to those skilled in the art, can be transcribed and/or translated to produce the mRNA for the polypeptide and/or a fragment thereof. The antisense strand is the complement of such a nucleic acid, and the encoding sequence can be deduced therefrom.

[0068] The terms "equivalent" or "biological equivalent" are used interchangeably when referring to a particular molecule, biological, or cellular material and intend those having minimal homology while still maintaining desired structure or functionality.

[0069] As used herein, the term "expression" refers to the process by which polynucleotides are transcribed into mRNA and/or the process by which the transcribed mRNA is subsequently being translated into peptides, polypeptides, or proteins. If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in a eukaryotic cell. The expression level of a gene may be determined by measuring the amount of mRNA or protein in a cell or tissue sample; further, the expression level of multiple genes can be determined to establish an expression profile for a particular sample.

[0070] As used herein, the term "sample" can refer to a composition comprising targets. Suitable samples for analysis by the disclosed methods, devices, and systems include cells, tissues, organs, or organisms or compositions obtained from cells, tissues or organisms. In some embodiments, samples are isolated from a subject.

[0071] As used herein, the term "functional" may be used to modify any molecule, biological, or cellular material to intend that it accomplishes a particular, specified effect.

[0072] A "gene delivery vehicle" is defined as any molecule that can carry inserted polynucleotides into a host cell. Examples of gene delivery vehicles are liposomes, micelles biocompatible polymers, including natural polymers and synthetic polymers; lipoproteins; polypeptides; polysaccharides; lipopolysaccharides; artificial viral envelopes; metal particles; and bacteria, or viruses, such as baculovirus, adenovirus and retrovirus, bacteriophage, cosmid, plasmid, fungal vectors and other recombination vehicles typically used in the art which have been described for expression in a variety of eukaryotic and prokaryotic hosts, and may be used for gene therapy as well as for simple protein expression.

[0073] A polynucleotide disclosed herein can be delivered to a cell or tissue using a gene delivery vehicle. "Gene delivery," "gene transfer," "transducing," and the like as used herein, are terms referring to the introduction of an exogenous polynucleotide (sometimes referred to as a "transgene") into a host cell, irrespective of the method used for the introduction. Such methods include a variety of well-known techniques such as vector-mediated gene transfer (by, e.g., viral infection/transfection, or various other protein-based or lipid-based gene delivery complexes) as well as techniques facilitating the delivery of "naked" polynucleotides (such as electroporation, "gene gun" delivery and various other techniques used for the introduction of polynucleotides). The introduced polynucleotide may be stably or transiently maintained in the host cell. Stable maintenance typically requires that the introduced polynucleotide either contains an origin of replication compatible with the host cell or integrates into a replicon of the host cell such as an extrachromosomal replicon (e.g., a plasmid) or a nuclear or mitochondrial chromosome. A number of "vectors" are known to be capable of mediating transfer of genes to mammalian cells, as is known in the art and described herein.

[0074] A "plasmid" is an extra-chromosomal DNA molecule separate from the chromosomal DNA which is capable of replicating independently of the chromosomal DNA. In many cases, it is circular and double-stranded. Plasmids provide a mechanism for horizontal gene transfer within a population of microbes and typically provide a selective advantage under a given environmental state. Plasmids may carry genes that provide resistance to naturally occurring antibiotics in a competitive environmental niche, or alternatively the proteins produced may act as toxins under similar circumstances.

[0075] "Plasmids" used in genetic engineering are called "plasmid vectors". Many plasmids are commercially available for such uses. The gene to be replicated is inserted into copies of a plasmid containing genes that make cells resistant to particular antibiotics and a multiple cloning site (MCS, or polylinker), which is a short region containing several commonly used restriction sites allowing the easy insertion of DNA fragments at this location. Another major use of plasmids is to make large amounts of proteins. In this case, researchers grow bacteria containing a plasmid harboring the gene of interest. Just as the bacterium produces proteins to confer its antibiotic resistance, it can also be induced to produce large amounts of proteins from the inserted gene.

[0076] A "yeast artificial chromosome" or "YAC" refers to a vector used to clone large DNA fragments (larger than 100 kb and up to 3000 kb). It is an artificially constructed chromosome and contains the telomeric, centromeric, and replication origin sequences needed for replication and preservation in yeast cells. Built using an initial circular plasmid, they are linearized by using restriction enzymes, and then DNA ligase can add a sequence or gene of interest within the linear molecule by the use of cohesive ends. Yeast expression vectors, such as YACs, YIps (yeast integrating plasmid), and YEps (yeast episomal plasmid), are extremely useful as one can get eukaryotic protein products with posttranslational modifications as yeasts are themselves eukaryotic cells, however YACs have been found to be more unstable than BACs, producing chimeric effects.

[0077] A "viral vector" is defined as a recombinantly produced virus or viral particle that comprises a polynucleotide to be delivered into a host cell, either in vivo, ex vivo or in vitro.

[0078] Examples of viral vectors include retroviral vectors, adenovirus vectors, adeno-associated virus vectors, alphavirus vectors and the like. Infectious tobacco mosaic virus (TMV)-based vectors can be used to manufacturer proteins and have been reported to express Griffithsin in tobacco leaves (O'Keefe et al. (2009) Proc. Nat. Acad. Sci. USA 106(15):6099-6104). Alphavirus vectors, such as Semliki Forest virus-based vectors and Sindbis virus-based vectors, have also been developed for use in gene therapy and immunotherapy. See, Schlesinger & Dubensky (1999) Curr. Opin. Biotechnol. 5:434-439 and Ying et al. (1999) Nat. Med. 5(7):823-827. In aspects where gene transfer is mediated by a retroviral vector, a vector construct refers to the polynucleotide comprising the retroviral genome or part thereof, and a therapeutic gene. Further details as to modern methods of vectors for use in gene

transfer may be found in, for example, Kotterman et al. (2015) Viral Vectors for Gene Therapy: Translational and Clinical Outlook Annual Review of Biomedical Engineering 17.

[0079] As used herein, "retroviral mediated gene transfer" or "retroviral transduction" carries the same meaning and refers to the process by which a gene or nucleic acid sequences are stably transferred into the host cell by virtue of the virus entering the cell and integrating its genome into the host cell genome. The virus can enter the host cell via its normal mechanism of infection or be modified such that it binds to a different host cell surface receptor or ligand to enter the cell. As used herein, retroviral vector refers to a viral particle capable of introducing exogenous nucleic acid into a cell through a viral or viral-like entry mechanism.

[0080] Retroviruses carry their genetic information in the form of RNA; however, once the virus infects a cell, the RNA is reverse-transcribed into the DNA form which integrates into the genomic DNA of the infected cell. The integrated DNA form is called a provirus.

[0081] In aspects where gene transfer is mediated by a DNA viral vector, such as an adenovirus (Ad) or adeno-associated virus (AAV), a vector construct refers to the polynucleotide comprising the viral genome or part thereof, and a transgene. Adenoviruses (Ads) are a relatively well characterized, homogenous group of viruses, including over 50 serotypes. Ads do not require integration into the host cell genome. Recombinant Ad derived vectors, particularly those that reduce the potential for recombination and generation of wild-type virus, have also been constructed. Such vectors are commercially available from sources such as Takara Bio USA (Mountain View, CA), Vector Biolabs (Philadelphia, PA), and Creative Biogene (Shirley, NY). Wild-type AAV has high infectivity and specificity integrating into the host cell's genome. See, Wold and Toth (2013) Curr. Gene. Ther. 13(6):421-433, Hermonat & Muzyczka (1984) Proc. Natl. Acad. Sci. USA 81:6466-6470, and Lebkowski et al. (1988) Mol. Cell. Biol. 8:3988-3996.

[0082] Vectors that contain both a promoter and a cloning site into which a polynucleotide can be operatively linked are well known in the art. Such vectors are capable of transcribing RNA *in vitro* or *in vivo*, and are commercially available from sources such as Agilent Technologies (Santa Clara, Calif.) and Promega Biotech (Madison, Wis.). In order to optimize expression and/or *in vitro* transcription, it may be necessary to remove, add or alter

5′ and/or 3′ untranslated portions of the clones to eliminate extra, potential inappropriate alternative translation initiation codons or other sequences that may interfere with or reduce expression, either at the level of transcription or translation. Alternatively, consensus ribosome binding sites can be inserted immediately 5′ of the start codon to enhance expression.

[0083] Gene delivery vehicles also include DNA/liposome complexes, micelles and targeted viral protein-DNA complexes. Liposomes that also comprise a targeting antibody or fragment thereof can be used in the methods disclosed herein. In addition to the delivery of polynucleotides to a cell or cell population, direct introduction of the proteins described herein to the cell or cell population can be done by the non-limiting technique of protein transfection, alternatively culturing conditions that can enhance the expression and/or promote the activity of the proteins disclosed herein are other non-limiting techniques.

[0084] "Homology" or "identity" or "similarity" refers to sequence similarity between two peptides or between two nucleic acid molecules. Homology can be determined by comparing a position in each sequence that may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same base or amino acid, then the molecules are homologous at that position. A degree of homology between sequences is a function of the number of matching or homologous positions shared by the sequences. An "unrelated" or "non-homologous" sequence shares less than 40% identity, or alternatively less than 25% identity, with one of the sequences of the present disclosure.

[0085] "Homology" or "identity" or "similarity" can also refer to two nucleic acid molecules that hybridize under stringent conditions.

[0086] "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 a PCR reaction, or the enzymatic cleavage of a polynucleotide by a ribozyme.

[0087] Examples of stringent hybridization conditions include: incubation temperatures of about 25° C. to about 37° C.; hybridization buffer concentrations of about 6×SSC to about 10×SSC; formamide concentrations of about 0% to about 25%; and wash solutions from about 4×SSC to about 8×SSC. Examples of moderate hybridization conditions include: incubation temperatures of about 40° C. to about 50° C.; buffer concentrations of about 9×SSC to about 2×SSC; formamide concentrations of about 30% to about 50%; and wash solutions of about 5×SSC to about 2×SSC. Examples of high stringency conditions include: incubation temperatures of about 55° C. to about 68° C.; buffer concentrations of about 1×SSC to about 0.1×SSC; formamide concentrations of about 55% to about 75%; and wash solutions of about 1×SSC, 0.1×SSC, or deionized water. In general, hybridization incubation times are from 5 minutes to 24 hours, with 1, 2, or more washing steps, and wash incubation times are about 1, 2, or 15 minutes. SSC is 0.15 M NaCl and 15 mM citrate buffer. It is understood that equivalents of SSC using other buffer systems can be employed.

[0088] As used herein, the term "specifically binds" refers to the binding specificity of a specific binding pair. Hybridization by a target-specific nucleic acid sequence of a particular target polynucleotide sequence in the presence of other potential targets is one characteristic of such binding. Specific binding involves two different nucleic acid molecules wherein one of the nucleic acid molecules specifically hybridizes with the second nucleic acid molecule through chemical or physical means. The two nucleic acid molecules are related in the sense that their binding with each other is such that they are capable of distinguishing their binding partner from other assay constituents having similar characteristics. The members of the binding component pair are referred to as ligand and receptor (anti-ligand), specific binding pair (SBP) member and SBP partner, and the like.

[0089] The term "isolated" as used herein refers to molecules or biologicals or cellular materials being substantially free from other materials.

[0090] As used herein, the term "linker" refers to a short peptide sequence that may occur between two protein domains. Linkers may often comprise flexible amino acid residues, *e.g.* glycine or serine, to allow for free movement of adjacent but fused protein domains. "XTEN" refers to any one of the exemplary linkers provided in Schellenberger et al. (2009) Nat Biotechnol. 27:1186–1190. doi: 10.1038/nbt.1588 or equivalent variants thereof.

[0091] As used herein, the term "organ" is a structure which is a specific portion of an individual organism, where a certain function or functions of the individual organism is locally performed and which is morphologically separate. Non-limiting examples of organs include the skin, blood vessels, cornea, thymus, kidney, heart, liver, umbilical cord, intestine, nerve, lung, placenta, pancreas, thyroid and brain.

[0092] The term"photospacer adjacent motif" or "PAM" refers to a sequence that activates the nuclease domain of Cas9. A "PAMmer" refers to a PAM-presenting oligonucleotide. As used herein, the term PAMmer generally refers to an antisense synthetic oligonucleotide composed alternating 2'OMe RNA and DNA bases and/or other variations of a PAM presenting oligonucleotide that can optimize the CRISPR/Cas9 system and generate specific cleavage of RNA targets without cross reactivity between non-target RNA or against genomic DNA. *See*, *e.g.*, O'Connell et al. (2014) Nature. 516(7530):263-266.

[0093] The term "promoter" as used herein refers to any sequence that regulates the expression of a coding sequence, such as a gene. Promoters may be constitutive, inducible, repressible, or tissue-specific, for example. A "promoter" is a control sequence that is a region of a polynucleotide sequence at which initiation and rate of transcription are controlled. It may contain genetic elements at which regulatory proteins and molecules may bind such as RNA polymerase and other transcription factors. Non-limiting exemplary promoters include CMV promoter and U6 promoter.

[0094] The term "protein", "peptide" and "polypeptide" are used interchangeably and in their broadest sense to refer to a compound of two or more subunits of amino acids, amino acid analogs or peptidomimetics. The subunits may be linked by peptide bonds. In another aspect, the subunit may be linked by other bonds, e.g., ester, ether, etc. A protein or peptide must contain at least two amino acids and no limitation is placed on the maximum number of amino acids which may comprise a protein's or peptide's sequence. Proteins and peptides are known to have a C-terminus, referring to the end with an unbound carboxy group on the terminal amino acid, and an N-terminus, referring to the end with an unbound amine group on the terminal amino acid. As used herein the term "amino acid" refers to either natural and/or unnatural or synthetic amino acids, including glycine and both the D and L optical isomers, amino acid analogs and peptidomimetics. The term "fused" in context of a protein

or polypeptide refers to the linkage between termini of two or more proteins or polypeptides (or domains thereof) to form a fusion protein.

[0095] As used herein, the term "recombinant expression system" refers to a genetic construct for the expression of certain genetic material or proteins formed by recombination.

[0096] As used herein, the term "subject" is used interchangeably with "patient" and is intended to mean any animal. In some embodiments, the subject may be a mammal. In some embodiments, the mammal is a non-human mammal. In some embodiments, the mammal is a bovine, equine, porcine, murine, feline, canine, simian, rat, or human.

[0097] The term "tissue" is used herein to refer to tissue of a living or deceased organism or any tissue derived from or designed to mimic a living or deceased organism. The tissue may be healthy, diseased, and/or have genetic mutations. The biological tissue may include any single tissue (e.g., a collection of cells that may be interconnected) or a group of tissues making up an organ or part or region of the body of an organism. The tissue may comprise a homogeneous cellular material or it may be a composite structure such as that found in regions of the body including the thorax which for instance can include lung tissue, skeletal tissue, and/or muscle tissue. Exemplary tissues include, but are not limited to those derived from liver, lung, thyroid, skin, pancreas, blood vessels, bladder, kidneys, brain, biliary tree, duodenum, abdominal aorta, iliac vein, heart and intestines, including any combination thereof.

[0098] As used herein, "treating" or "treatment" of a disease in a subject refers to (1) preventing the symptoms or disease from occurring in a subject that is predisposed or does not yet display symptoms of the disease; (2) inhibiting the disease or arresting its development; or (3) ameliorating or causing regression of the disease or the symptoms of the disease. As understood in the art, "treatment" is an approach for obtaining beneficial or desired results, including clinical results. For the purposes of the present technology, beneficial or desired results can include one or more, but are not limited to, alleviation or amelioration of one or more symptoms, diminishment of extent of a condition (including a disease), stabilized (*i.e.*, not worsening) state of a condition (including disease), delay or slowing of condition (including disease), progression, amelioration or palliation of the

condition (including disease), states and remission (whether partial or total), whether detectable or undetectable.

[0099] As used herein, the term "vector" intends a recombinant vector that retains the ability to infect and transduce non-dividing and/or slowly-dividing cells and integrate into the target cell's genome. The vector may be derived from or based on a wild-type virus. Aspects of this disclosure relate to an adeno-associated virus vector.

[0100] A number of other vector elements are disclosed herein; e.g., plasmids, promoters, linkers, signals, etc. The nature and function of these vector elements are commonly understood in the art and a number of these vector elements are commercially available. Non-limiting exemplary sequences thereof, e.g., SEQ ID NOS: 1-8 are disclosed herein and further description thereof is provided herein below and/or illustrated in FIGs. 3-10.

CRISPR/Cas directed RNA-editing (CREDIT)

[0101] Disclosed herein is an efficient, versatile and simplified platform technology for performing programmable RNA editing at single-nucleotide resolution using RNA-targeting CRISPR/Cas (RCas). This approach, which Applicants have termed "Cas-directed RNA editing" or "CREDIT," provides a means to reversibly alter genetic information in a temporal manner, unlike traditional CRISPR/Cas9 driven genomic engineering which relies on permanently altering DNA sequence. Recombinant expression systems are engineered to induce edits to specific RNA bases as determined by the guide RNA design. As such, in some embodiments, Applicants provide a fully encodeable recombinant expression system comprising a nuclease-dead version of Streptococcus pyogenes Cas9 (dCas9) fused to an ADAR deaminase domain and a corresponding extended single guide RNA (esgRNA). In some embodiments, the system generates recombinant proteins with effector deaminase enzyme complexes capable of performing ribonucleotide base modification to alter how the sequence of the RNA molecule is recognized by cellular machinery. In some embodiments, the CREDIT expression system comprises A) a nucleic acid sequence encoding a nuclease-dead CRISPR associated endonuclease (dCas) fused to a catalytically active deaminase domain of ADAR (Adenosine Deaminase acting on RNA) and B) an extended single guide RNA (esgRNA) sequence comprising i) a short extension sequence of homology to the target RNA comprising a mismatch for a target adenosine, ii) a dCas scaffold binding sequence, and

optionally iii) a sequence complementary to the target RNA sequence (also known as a spacer sequence in a sgRNA context). Exemplary constructs that express CREDIT expression system components include, without limitation, dCas9 fused to catalytically active deaminase domains of human ADAR2 (hADAR2DD, E488QhADAR2DD) using an 'XTEN' linker peptide for spatial separation (FIG. 1B). With dCas9 as a surrogate RBD (RNA-Binding Domain), Applicants engineered and customized single guide RNAs (sgRNAs) with unique short extension sequences (esgRNA) to direct hADAR2DD to RNA sites for target specific A – I editing. For the purposes of the present disclosure, CRISPR/Cas associated endonucleases other than Cas9 or Cas9 orthologs (e.g., Cas13 (also known as C2c2), Cpf1, Cas6f/Csy4, CasX, CasY, and CasRx) are also provided herein for use in the CREDIT expression system. See also Wright et al., Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering, Cell, Vol. 164 (1-2): 29-44, 2016.

[0102] In some embodiments disclosed herein, dCas polypeptide has been engineered to recognize a target RNA, wherein the inactive Cas polypeptide is associated with an effector. In some embodiments, the dCas polypeptide is a *Streptococcus pyogenes d*Cas9 polypeptide. In some embodiments, the dCas9 polypeptide comprises a mutation, such as D10A, H840A, or both, in the *Streptococcus pyogenes* Cas9 polypeptide. This repurposed or engineered dCas9 polypeptide-comprising nucleoprotein complex that binds to RNA is referred to herein as RdCas9. CRISPR has revolutionized genome engineering by allowing simply-programmed recognition of DNA in human cells and supported related technologies in imaging and gene expression modulation. In WO 2017/091630, incorporated by reference in its entirety herein, an analogous means to target RNA using an RCas9 was developed. In this earlier work, engineered nucleoprotein complexes comprise a Cas9 protein and a single guide RNA (sgRNA). Together, the Cas9 protein and sgRNA components were engineered to hypothetically recognize any target RNA sequence. Optionally, in such systems, an (chemically-modified or synthetic) antisense PAMmer oligonucleotide could be included in the RCas9 system to simulate a DNA substrate for recognition by Cas9 via hybridization to the target RNA. However, surprisingly highly effective RNA targeting without PAMmer was also shown. Now, herein is disclosed RdCas-ADAR RNA editing systems which do not require a PAMmer and as such are fully encodeable Cas9-mediated RNA targeting systems which provide a reversible platform for modification of target RNA.

[0103] For the purposes of the present disclosure, Cas9 endonucleases used herein include, without limitation, orthologs derived from archaeal or bacterial Cas9 polypeptides. Such polypeptides can be derived from, without limitations *Haloferax mediteranii*, Mycobacterium *tuberculosis, Francisella tularensis subsp. novicida, Pasteurella multocida, Neisseria meningitidis, Campylobacter jejune, Streptococcus thermophilus* LMD-9 CRISPR 3, *Campylobacter lari* CF89- 12, *Mycoplasma gallisepticum str. F, Nitratifractor salsuginis str* DSM 1651 1, *Parvibaculum lavamentivorans, Roseburia intestinalis, Neisseria cinerea, Gluconacetobacter diazotrophicus, Azospirillum* B510, *Sphaerochaeta globus str. Buddy, Flavobacterium columnare, Fluviicola taffensis, Bacteroides coprophilus, Mycoplasma mobile, Lactobacillus farciminis, Streptococcus pasteurianus, Lactobacillus johnsonii, Staphylococcus pseudintermedius, Filifactor alocis, Treponema denticola, Legionella pneumophila str. Paris, Sutterella wadsworthensis, Corynebacter diphtheriae, or Streptococcus aureus; Francisella novicida (e.g., Francisella novicida* CPf1), or *Natronobacterium gregoryi* Argonaute. Each of these respective candidate Cas polypeptides are modified and/or repurposed to target RNA and fused to an ADAR deaminase domain for use in the systems disclosed herein, which system additionally comprises an extended sgRNA (esgRNA) which comprises a guide "scaffold sequence" which comprises all or part of, or is derived from, the wild type (WT) cognate guide nucleic acid of each of these respective bacteria or archaeal organisms. In some embodiments, Cas endonucleases for use herein include, without limitation, Cas13 (c2C2), Cpf1, CasX, CasY, and CasRx.

[0104] Further nonlimiting examples of orthologs and biological equivalents Cas9 are provided in the table below:

Name	**Protein Sequence**
S. pyogenes Cas9 SEQ ID NO: 1	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD SGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGE QKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLK RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKE

	DIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIV IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLY YLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG FDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGY KEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYN KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQS ITGLYETRIDLSQLGGD*
Staphylococcus aureus Cas9 SEQ ID NO: 2	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGAR RLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAA LLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKD GEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGP GEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVI TRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPE FTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEI EQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKE IPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINE MQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLN NPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYET FKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLM NLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIAN ADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIK DFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKL KKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYS KKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGV YKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKING ELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYST DILGNLYEVKSKKHPQIIKKG*

S. thermophilus CRISPR 1 Cas9 SEQ ID NO: 3	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRL ARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKN MVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLERYQTYGQ LRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEILT GKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFRAAKASYT AQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLS CDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTLN TEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELI PELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKI VNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLK AANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSN QFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFREL KAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQE HFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNL WKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAPYQHFVDTLKSKEFEDSI LFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYTQDGYDAF MKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINDKGKEVPCNPFLKYKE EHGYIRKYSKKGNGPEIKSLKYYDSKLGNHIDITPKDSNNKVVLQSVSPWRADV YFNKTTGKYEILGLKYADLQFDKGTGTYKISQEKYNDIKKKEGVDSDSEFKFTL YKNDLLLVKDTETKEQQLFRFLSRTMPKQKHYVELKPYDKQKFEGGEALIKVL GNVANSGQCKKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPKLDF*
N. meningitidis Cas9 SEQ ID NO: 4	MAAFKPNPINYILGLDIGIASVGWAMVEIDEDENPICLIDLGVRVFERAEVPKTG DSLAMARRLARSVRRLTRRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPN TPWQLRAAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKG VADNAHALQTGDFRTPAELALNKFEKESGHIRNQRGDYSHTFSRKDLQAELILL FEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDAVQKMLGHCTFEPAEPKAA KNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSKLTYAQARK LLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNLS PELQDEIGTAFSLFKTDEDITGRLKDRIQPEILEALLKHISFDKFVQISLKALRRIV PLMEQGKRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQARK VINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENRKDREKAAAKFREY FPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLGRLNEKGYVEIDHALPFSRT WDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSRFPRS KKQRILLQKFDEDGFKERNLNDTRYVNRFLCQFVADRMRLTGKGKKRVFASN GQITNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRYKEMN AFDGKTIDKETGEVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTPEK LRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGQGHMETVKSAKRLDEGVS VLRVPLTQLKLKDLEKMVNREREPKLYEALKARLEAHKDDPAKAFAEPFYKY DKAGNRTQQVKAVRVEQVQKTGVWVRNHNGIADNATMVRVDVFEKGDKYY LVPIYSWQVAKGILPDRAVVQGKDEEDWQLIDDSFNFKFSLHPNDLVEVITKKA RMFGYFASCHRGTGNINIRIHDLDHKIGKNGILEGIGVKTALSFQKYQIDELGKEI RPCRLKKRPPVR*

Parvibaculum lavamentivorans Cas9 SEQ ID NO: 5	MERIFGFDIGTTSIGFSVIDYSSTQSAGNIQRLGVRIFPEARDPDGTPLNQQRRQK RMMRRQLRRRRIRRKALNETLHEAGFLPAYGSADWPVVMADEPYELRRRGLE EGLSAYEFGRAIYHLAQHRHFKGRELEESDTPDPDVDDEKEAANERAATLKAL KNEQTTLGAWLARRPPSDRKRGIHAHRNVVAEEFERLWEVQSKFHPALKSEEM RARISDTIFAQRPVFWRKNTLGECRFMPGEPLCPKGSWLSQQRRMLEKLNNLAI AGGNARPLDAEERDAILSKLQQQASMSWPGVRSALKALYKQRGEPGAEKSLK FNLELGGESKLLGNALEAKLADMFGPDWPAHPRKQEIRHAVHERLWAADYGE TPDKKRVIILSEKDRKAHREAAANSFVADFGITGEQAAQLQALKLPTGWEPYSI PALNLFLAELEKGERFGALVNGPDWEGWRRTNFPHRNQPTGEILDKLPSPASKE ERERISQLRNPTVVRTQNELRKVVNNLIGLYGKPDRIRIEVGRDVGKSKREREEI QSGIRRNEKQRKKATEDLIKNGIANPSRDDVEKWILWKEGQERCPYTGDQIGFN ALFREGRYEVEHIWPRSRSFDNSPRNKTLCRKDVNIEKGNRMPFEAFGHDEDR WSAIQIRLQGMVSAKGGTGMSPGKVKRFLAKTMPEDFAARQLNDTRYAAKQI LAQLKRLWPDMGPEAPVKVEAVTGQVTAQLRKLWTLNNILADDGEKTRADH RHHAIDALTVACTHPGMTNKLSRYWQLRDDPRAEKPALTPPWDTIRADAEKA VSEIVVSHRVRKKVSGPLHKETTYGDTGTDIKTKSGTYRQFVTRKKIESLSKGEL DEIRDPRIKEIVAAHVAGRGGDPKKAFPPYPCVSPGGPEIRKVRLTSKQQLNLM AQTGNGYADLGSNHHIAIYRLPDGKADFEIVSLFDASRRLAQRNPIVQRTRADG ASFVMSLAAGEAIMIPEGSKKGIWIVQGVWASGQVVLERDTDADHSTTTRPMP NPILKDDAKKVSIDPIGRVRPSND*
Corynebacter diphtheria Cas9 SEQ ID NO: 6	MKYHVGIDVGTFSVGLAAIEVDDAGMPIKTLSLVSHIHDSGLDPDEIKSAVTRL ASSGIARRTRRLYRRKRRRLQQLDKFIQRQGWPVIELEDYSDPLYPWKVRAELA ASYIADEKERGEKLSVALRHIARHRGWRNPYAKVSSLYLPDGPSDAFKAIREEI KRASGQPVPETATVGQMVTLCELGTLKLRGEGGVLSARLQQSDYAREIQEICR MQEIGQELYRKIIDVVFAAESPKGSASSRVGKDPLQPGKNRALKASDAFQRYRI AALIGNLRVRVDGEKRILSVEEKNLVFDHLVNLTPKKEPEWVTIAEILGIDRGQL IGTATMTDDGERAGARPPTHDTNRSIVNSRIAPLVDWWKTASALEQHAMVKAL SNAEVDDFDSPEGAKVQAFFADLDDDVHAKLDSLHLPVGRAAYSEDTLVRLTR RMLSDGVDLYTARLQEFGIEPSWTPPTPRIGEPVGNPAVDRVLKTVSRWLESAT KTWGAPERVIIEHVREGFVTEKRAREMDGDMRRRAARNAKLFQEMQEKLNVQ GKPSRADLWRYQSVQRQNCQCAYCGSPITFSNSEMDHIVPRAGQGSTNTRENL VAVCHRCNQSKGNTPFAIWAKNTSIEGVSVKEAVERTRHWVTDTGMRSTDFK KFTKAVVERFQRATMDEEIDARSMESVAWMANELRSRVAQHFASHGTTVRVY RGSLTAEARRASGISGKLKFFDGVGKSRLDRRHHAIDAAVIAFTSDYVAETLAV RSNLKQSQAHRQEAPQWREFTGKDAEHRAAWRVWCQKMEKLSALLTEDLRD DRVVVMSNVRLRLGNGSAHKETIGKLSKVKLSSQLSVSDIDKASSEALWCALT REPGFDPKEGLPANPERHIRVNGTHVYAGDNIGLFPVSAGSIALRGGYAELGSSF HHARVYKITSGKKPAFAMLRVYTIDLLPYRNQDLFSVELKPQTMSMRQAEKKL RDALATGNAEYLGWLVVDDELVVDTSKIATDQVKAVEAELGTIRRWRVDGFF SPSKLRLRPLQMSKEGIKKESAPELSKIIDRPGWLPAVNKLFSDGNVTVVRRDSL GRVRLESTAHLPVTWKVQ*

Streptococcus pasteurianus Cas9 SEQ ID NO: 7	MTNGKILGLDIGIASVGVGIIEAKTGKVVHANSRLFSAANAENNAERRGFRGSR RLNRRKKHRVKRVRDLFEKYGIVTDFRNLNLNPYELRVKGLTEQLKNEELFAA LRTISKRRGISYLDDAEDDSTGSTDYAKSIDENRRLLKNKTPGQIQLERLEKYGQ LRGNFTVYDENGEAHRLINVFSTSDYEKEARKILETQADYNKKITAEFIDDYVEI LTQKRKYYHGPGNEKSRTDYGRFRTDGTTLENIFGILIGKCNFYPDEYRASKAS YTAQEYNFLNDLNNLKVSTETGKLSTEQKESLVEFAKNTATLGPAKLLKEIAKI LDCKVDEIKGYREDDKGKPDLHTFEPYRKLKFNLESINIDDLSREVIDKLADILT LNTEREGIEDAIKRNLPNQFTEEQISEIIKVRKSQSTAFNKGWHSFSAKLMNELIP ELYATSDEQMTILTRLEKFKVNKKSSKNTKTIDEKEVTDEIYNPVVAKSVRQTIK IINAAVKKYGDFDKIVIEMPRDKNADDEKKFIDKRNKENKKEKDDALKRAAYL YNSSDKLPDEVFHGNKQLETKIRLWYQQGERCLYSGKPISIQELVHNSNNFEID HILPLSLSFDDSLANKVLVYAWTNQEKGQKTPYQVIDSMDAAWSFREMKDYV LKQKGLGKKKRDYLLTTENIDKIEVKKKFIERNLVDTRYASRVVLNSLQSALRE LGKDTKVSVVRGQFTSQLRRKWKIDKSRETYHHHAVDALIIAASSQLKLWEKQ DNPMFVDYGKNQVVDKQTGEILSVSDDEYKELVFQPPYQGFVNTISSKGFEDEI LFSYQVDSKYNRKVSDATIYSTRKAKIGKDKKEETYVLGKIKDIYSQNGFDTFIK KYNKDKTQFLMYQKDSLTWENVIEVILRDYPTTKKSEDGKNDVKCNPFEEYRR ENGLICKYSKKGKGTPIKSLKYYDKKLGNCIDITPEESRNKVILQSINPWRADVY FNPETLKYELMGLKYSDLSFEKGTGNYHISQEKYDAIKEKEGIGKKSEFKFTLY RNDLILIKDIASGEQEIYRFLSRTMPNVNHYVELKPYDKEKFDNVQELVEALGE ADKVGRCIKGLNKPNISIYKVRTDVLGNKYFVKKKGDKPKLDFKNNKK*
Neisseria cinerea Cas9 SEQ ID NO: 8	MAAFKPNPMNYILGLDIGIASVGWAIVEIDEEENPIRLIDLGVRVFERAEVPKTG DSLAAARRLARSVRRLTRRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPN TPWQLRAAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKG VADNTHALQTGDFRTPAELALNKFEKESGHIRNQRGDYSHTFNRKDLQAELNL LFEKQKEFGNPHVSDGLKEGIETLLMTQRPALSGDAVQKMLGHCTFEPTEPKA AKNTYTAERFVWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSKLTYAQA RKLLDLDDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPL NLSPELQDEIGTAFSLFKTDEDITGRLKDRVQPEILEALLKHISFDKFVQISLKAL RRIVPLMEQGNRYDEACTEIYGDHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQ ARKVINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENRKDREKSAAKF REYFPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLGRLNEKGYVEIDHALP FSRTWDDSFNNKVLALGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSR FPRSKKQRILLQKFDEDGFKERNLNDTRYINRFLCQFVADHMLLTGKGKRRVF ASNGQITNLLRGFWGLRKVRAENDRHHALDAVVVACSTIAMQQKITRFVRYKE MNAFDGKTIDKETGEVLHQKAHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADT PEKLRTLLAEKLSSRPEAVHKYVTPLFISRAPNRKMSGQGHMETVKSAKRLDE GISVLRVPLTQLKLKDLEKMVNREREPKLYEALKARLEAHKDDPAKAFAEPFY KYDKAGNRTQQVKAVRVEQVQKTGVWVHNHNGIADNATIVRVDVFEKGGKY YLVPIYSWQVAKGILPDRAVVQGKDEEDWTVMDDSFEFKFVLYANDLIKLTAK KNEFLGYFVSLNRATGAIDIRTHDTDSTKGKNGIFQSVGVKTALSFQKYQIDEL GKEIRPCRLKKRPPVR*

Campylobacter lari Cas9 SEQ ID NO: 9	MRILGFDIGINSIGWAFVENDELKDCGVRIFTKAENPKNKESLALPRRNARSSRR RLKRRKARLIAIKRILAKELKLNYKDYVAADGELPKAYEGSLASVYELRYKALT QNLETKDLARVILHIAKHRGYMNKNEKKSNDAKKGKILSALKNNALKLENYQS VGEYFYKEFFQKYKKNTKNFIKIRNTKDNYNNCVLSSDLEKELKLILEKQKEFG YNYSEDFINEILKVAFFQRPLKDFSHLVGACTFFEEEKRACKNSYSAWEFVALT KIINEIKSLEKISGEIVPTQTINEVLNLILDKGSITYKKFRSCINLHESISFKSLKYDK ENAENAKLIDFRKLVEFKKALGVHSLSRQELDQISTHITLIKDNVKLKTVLEKYN LSNEQINNLLEIEFNDYINLSFKALGMILPLMREGKRYDEACEIANLKPKTVDEK KDFLPAFCDSIFAHELSNPVVNRAISEYRKVLNALLKKYGKVHKIHLELARDVG LSKKAREKIEKEQKENQAVNAWALKECENIGLKASAKNILKLKLWKEQKEICIY SGNKISIEHLKDEKALEVDHIYPYSRSFDDSFINKVLVFTKENQEKLNKTPFEAF GKNIEKWSKIQTLAQNLPYKKKNKILDENFKDKQQEDFISRNLNDTRYIATLIAK YTKEYLNFLLLSENENANLKSGEKGSKIHVQTISGMLTSVLRHTWGFDKKDRN NHLHHALDAIIVAYSTNSIIKAFSDFRKNQELLKARFYAKELTSDNYKHQVKFFE PFKSFREKILSKIDEIFVSKPPRKRARRALHKDTFHSENKIIDKCSYNSKEGLQIAL SCGRVRKIGTKYVENDTIVRVDIFKKQNKFYAIPIYAMDFALGILPNKIVITGKD KNNNPKQWQTIDESYEFCFSLYKNDLILLQKKNMQEPEFAYYNDFSISTSSICVE KHDNKFENLTSNQKLLFSNAKEGSVKVESLGIQNLKVFEKYIITPLGDKIKADFQ PRENISLKTSKKYGLR*
T. denticola Cas9 SEQ ID NO: 10	MKKEIKDYFLGLDVGTGSVGWAVTDTDYKLLKANRKDLWGMRCFETAETAE VRRLHRGARRRIERRKKRIKLLQELFSQEIAKTDEGFFQRMKESPFYAEDKTILQ ENTLFNDKDFADKTYHKAYPTINHLIKAWIENKVKPDPRLLYLACHNIIKKRGH FLFEGDFDSENQFDTSIQALFEYLREDMEVDIDADSQKVKEILKDSSLKNSEKQS RLNKILGLKPSDKQKKAITNLISGNKINFADLYDNPDLKDAEKNSISFSKDDFDA LSDDLASILGDSFELLLKAKAVYNCSVLSKVIGDEQYLSFAKVKIYEKHKTDLT KLKNVIKKHFPKDYKKVFGYNKNEKNNNNYSGYVGVCKTKSKKLIINNSVNQ EDFYKFLKTILSAKSEIKEVNDILTEIETGTFLPKQISKSNAEIPYQLRKMELEKIL SNAEKHFSFLKQKDEKGLSHSEKIIMLLTFKIPYYIGPINDNHKKFFPDRCWVVK KEKSPSGKTTPWNFFDHIDKEKTAEAFITSRTNFCTYLVGESVLPKSSLLYSEYT VLNEINNLQIIIDGKNICDIKLKQKIYEDLFKKYKKITQKQISTFIKHEGICNKTDE VIILGIDKECTSSLKSYIELKNIFGKQVDEISTKNMLEEIIRWATIYDEGEGKTILK TKIKAEYGKYCSDEQIKKILNLKFSGWGRLSRKFLETVTSEMPGFSEPVNIITAM RETQNNLMELLSSEFTFTENIKKINSGFEDAEKQFSYDGLVKPLFLSPSVKKML WQTLKLVKEISHITQAPPKKIFIEMAKGAELEPARTKTRLKILQDLYNNCKNDA DAFSSEIKDLSGKIENEDNLRLRSDKLYLYYTQLGKCMYCGKPIEIGHVFDTSNY DIDHIYPQSKIKDDSISNRVLVCSSCNKNKEDKYPLKSEIQSKQRGFWNFLQRNN FISLEKLNRLTRATPISDDETAKFIARQLVETRQATKVAAKVLEKMFPETKIVYS KAETVSMFRNKFDIVKCREINDFHHAHDAYLNIVVGNVYNTKFTNNPWNFIKE KRDNPKIADTYNYYKVFDYDVKRNNITAWEKGKTIITVKDMLKRNTPIYTRQA ACKKGELFNQTIMKKGLGQHPLKKEGPFSNISKYGGYNKVSAAYYTLIEYEEK GNKIRSLETIPLYLVKDIQKDQDVLKSYLTDLLGKKEFKILVPKIKINSLLKINGF PCHITGKTNDSFLLRPAVQFCCSNNEVLYFKKIIRFSEIRSQREKIGKTISPYEDLS FRSYIKENLWKKTKNDEIGEKEFYDLLQKKNLEIYDMLLTKHKDTIYKKRPNSA TIDILVKGKEKFKSLIIENQFEVILEILKLFSATRNVSDLQHIGGSKYSGVAKIGNK ISSLDNCILIYQSITGIFEKRIDLLKV*
S. mutans Cas9 SEQ ID NO: 11	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIEKNLLGALL FDSGNTAEDRRLKRTARRRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFL VTEDKRGERHPIFGNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAH IIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKI SKSAKKDRVLKLFPNEKSNGRFAEFLKLIVGNQADFKKHFELEEKAPLQFSKDT YEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVGTKAPLSASMIQRY NEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKDGYAGYIDGKTNQEAFYKYLK GLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFL ADNQDRIEKLLTFRIPYYVGPLARGKSDFAWLSRKSADKITPWNFDEIVDKESS

	AEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFD ANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKVFNASY GTYHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQ VKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDA LSFKEEIAKAQVIGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPE NIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRL FLYYLQNGRDMYTGEELDIDYLSQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGK SDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKR QLVETRQITKHVARILDERFNTETDENNKKIRQVKIVTLKSNLVSNFRKEFELYK VREINDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATA KKFFYSNIMNFFKKDDVRTDKNGEIIWKKDEHISNIKKVLSYPQVNIVKKVEEQ TGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKS KKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEF KELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAP ATFKFFDKNIDRKRYTSTTEILNATLIHQSITGLYETRIDLNKLGGD
S. thermophilus CRISPR 3 Cas9 SEQ ID NO: 12	MTKPYSIGLDIGTNSVGWAVTTDNYKVPSKKMKVLGNTSKKYIKKNLLGVLLF DSGITAEGRRLKRTARRRYTRRRNRILYLQEIFSTEMATLDDAFFQRLDDSFLVP DDKRDSKYPIFGNLVEEKAYHDEFPTIYHLRKYLADSTKKADLRLVYLALAHM IKYRGHFLIEGEFNSKNNDIQKNFQDFLDTYNAIFESDLSLENSKQLEEIVKDKIS KLEKKDRILKLFPGEKNSGIFSEFLKLIVGNQADFRKCFNLDEKASLHFSKESYD EDLETLLGYIGDDYSDVFLKAKKLYDAILLSGFLTVTDNETEAPLSSAMIKRYN EHKEDLALLKEYIRNISLKTYNEVFKDDTKNGYAGYIDGKTNQEDFYVYLKKL LAEFEGADYFLEKIDREDFLRKQRTFDNGSIPYQIHLQEMRAILDKQAKFYPFLA KNKERIEKILTFRIPYYVGPLARGNSDFAWSIRKRNEKITPWNFEDVIDKESSAE AFINRMTSFDLYLPEEKVLPKHSLLYETFNVYNELTKVRFIAESMRDYQFLDSK QKKDIVRLYFKDKRKVTDKDIIEYLHAIYGYDGIELKGIEKQFNSSLSTYHDLLN IINDKEFLDDSSNEAIIEEIIHTLTIFEDREMIKQRLSKFENIFDKSVLKKLSRRHYT GWGKLSAKLINGIRDEKSGNTILDYLIDDGISNRNFMQLIHDDALSFKKKIQKAQ IIGDEDKGNIKEVVKSLPGSPAIKKGILQSIKIVDELVKVMGGRKPESIVVEMARE NQYTNQGKSNSQQRLKRLEKSLKELGSKILKENIPAKLSKIDNNALQNDRLYLY YLQNGKDMYTGDDLDIDRLSNYDIDHIIPQAFLKDNSIDNKVLVSSASNRGKSD DVPSLEVVKKRKTFWYQLLKSKLISQRKFDNLTKAERGGLSPEDKAGFIQRQLV ETRQITKHVARLLDEKFNNKKDENNRAVRTVKIITLKSTLVSQFRKDFELYKVR EINDFHHAHDAYLNAVVASALLKKYPKLEPEFVYGDYPKYNSFRERKSATEKV YFYSNIMNIFKKSISLADGRVIERPLIEVNEETGESVWNKESDLATVRRVLSYPQ VNVVKKVEEQNHGLDRGKPKGLFNANLSSKPKPNSNENLVGAKEYLDPKKYG GYAGISNSFTVLVKGTIEKGAKKKITNVLEFQGISILDRINYRKDKLNFLLEKGY KDIELIIELPKYSLFELSDGSRRMLASILSTNNKRGEIHKGNQIFLSQKFVKLLYH AKRISNTINENHRKYVENHKKEFEELFYYILEFNENYVGAKKNGKLLNSAFQSW QNHSIDELCSSFIGPTGSERKGLFELTSRGSAADFEFLGVKIPRYRDYTPSSLLKD ATLIHQSVTGLYETRIDLAKLGEG

C. jejuni Cas9 SEQ ID NO: 13	MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLARSAR KRLARRKARLNHLKHLIANEFKLNYEDYQSFDESLAKAYKGSLISPYELRFRAL NELLSKQDFARVILHIAKRRGYDDIKNSDDKEKGAILKAIKQNEEKLANYQSVG EYLYKEYFQKFKENSKEFTNVRNKKESYERCIAQSFLKDELKLIFKKQREFGFSF SKKFEEEVLSVAFYKRALKDFSHLVGNCSFFTDEKRAPKNSPLAFMFVALTRIIN LLNNLKNTEGILYTKDDLNALLNEVLKNGTLTYKQTKKLLGLSDDYEFKGEKG TYFIEFKKYKEFIKALGEHNLSQDDLNEIAKDITLIKDEIKLKKALAKYDLNQNQ IDSLSKLEFKDHLNISFKALKLVTPLMLEGKKYDEACNELNLKVAINEDKKDFL PAFNETYYKDEVTNPVVLRAIKEYRKVLNALLKKYGKVHKINIELAREVGKNH SQRAKIEKEQNENYKAKKDAELECEKLGLKINSKNILKLRLFKEQKEFCAYSGE KIKISDLQDEKMLEIDHIYPYSRSFDDSYMNKVLVFTKQNQEKLNQTPFEAFGN DSAKWQKIEVLAKNLPTKKQKRILDKNYKDKEQKNFKDRNLNDTRYIARLVL NYTKDYLDFLPLSDDENTKLNDTQKGSKVHVEAKSGMLTSALRHTWGFSAKD RNNHLHHAIDAVIIAYANNSIVKAFSDFKKEQESNSAELYAKKISELDYKNKRK FFEPFSGFRQKVLDKIDEIFVSKPERKKPSGALHEETFRKEEEFYQSYGGKEGVL KALELGKIRKVNGKIVKNGDMFRVDIFKHKKTNKFYAVPIYTMDFALKVLPNK AVARSKKGEIKDWILMDENYEFCFSLYKDSLILIQTKDMQEPEFVYYNAFTSST VSLIVSKHDNKFETLSKNQKILFKNANEKEVIAKSIGIQNLKVFEKYIVSALGEVT KAEFRQREDFKK
P. multocida Cas9 SEQ ID NO: 14	MQTTNLSYILGLDLGIASVGWAVVEINENEDPIGLIDVGVRIFERAEVPKTGESL ALSRRLARSTRRLIRRRAHRLLLAKRFLKREGILSTIDLEKGLPNQAWELRVAGL ERRLSAIEWGAVLLHLIKHRGYLSKRKNESQTNNKELGALLSGVAQNHQLLQS DDYRTPAELALKKFAKEEGHIRNQRGAYTHTFNRLDLLAELNLLFAQQHQFGN PHCKEHIQQYMTELLMWQKPALSGEAILKMLGKCTHEKNEFKAAKHTYSAER FVWLTKLNNLRILEDGAERALNEEERQLLINHPYEKSKLTYAQVRKLLGLSEQA IFKHLRYSKENAESATFMELKAWHAIRKALENQGLKDTWQDLAKKPDLLDEIG TAFSLYKTDEDIQQYLTNKVPNSVINALLVSLNFDKFIELSLKSLRKILPLMEQG KRYDQACREIYGHHYGEANQKTSQLLPAIPAQEIRNPVVLRTLSQARKVINAIIR QYGSPARVHIETGRELGKSFKERREIQKQQEDNRTKRESAVQKFKELFSDFSSEP KSKDILKFRLYEQQHGKCLYSGKEINIHRLNEKGYVEIDHALPFSRTWDDSFNN KVLVLASENQNKGNQTPYEWLQGKINSERWKNFVALVLGSQCSAAKKQRLLT QVIDDNKFIDRNLNDTRYIARFLSNYIQENLLLVGKNKKNVFTPNGQITALLRSR WGLIKARENNNRHHALDAIVVACATPSMQQKITRFIRFKEVHPYKIENRYEMV DQESGEIISPHFPEPWAYFRQEVNIRVFDNHPDTVLKEMLPDRPQANHQFVQPL FVSRAPTRKMSGQGHMETIKSAKRLAEGISVLRIPLTQLKPNLLENMVNKEREP ALYAGLKARLAEFNQDPAKAFATPFYKQGGQQVKAIRVEQVQKSGVLVRENN GVADNASIVRTDVFIKNNKFFLVPIYTWQVAKGILPNKAIVAHKNEDEWEEMD EGAKFKFSLFPNDLVELKTKKEYFFGYYIGLDRATGNISLKEHDGEISKGKDGV YRVGVKLALSFEKYQVDELGKNRQICRPQQRQPVR
F. novicida Cas9 SEQ ID NO: 15	MNFKILPIAIDLGVKNTGVFSAFYQKGTSLERLDNKNGKVYELSKDSYTLLMNN RTARRHQRRGIDRKQLVKRLFKLIWTEQLNLEWDKDTQQAISFLFNRRGFSFIT DGYSPEYLNIVPEQVKAILMDIFDDYNGEDDLDSYLKLATEQESKISEIYNKLM QKILEFKLMKLCTDIKDDKVSTKTLKEITSYEFELLADYLANYSESLKTQKFSYT DKQGNLKELSYYHHDKYNIQEFLKRHATINDRILDTLLTDDLDIWNFNFEKFDF DKNEEKLQNQEDKDHIQAHLHHFVFAVNKIKSEMASGGRHRSQYFQEITNVLD ENNHQEGYLKNFCENLHNKKYSNLSVKNLVNLIGNLSNLELKPLRKYFNDKIH AKADHWDEQKFTETYCHWILGEWRVGVKDQDKKDGAKYSYKDLCNELKQK VTKAGLVDFLLELDPCRTIPPYLDNNNRKPPKCQSLILNPKFLDNQYPNWQQYL QELKKLQSIQNYLDSFETDLKVLKSSKDQPYFVEYKSSNQQIASGQRDYKDLDA RILQFIFDRVKASDELLLNEIYFQAKKLKQKASSELEKLESSKKLDEVIANSQLSQ ILKSQHTNGIFEQGTFLHLVCKYYKQRQRARDSRLYIMPEYRYDKKLHKYNNT GRFDDDNQLLTYCNHKPRQKRYQLLNDLAGVLQVSPNFLKDKIGSDDDLFISK WLVEHIRGFKKACEDSLKIQKDNRGLLNHKINIARNTKGKCEKEIFNLICKIEGS

	EDKKGNYKHGLAYELGVLLFGEPNEASKPEFDRKIKKFNSIYSFAQIQQIAFAER KGNANTCAVCSADNAHRMQQIKITEPVEDNKDKIILSAKAQRLPAIPTRIVDGA VKKMATILAKNIVDDNWQNIKQVLSAKHQLHIPIITESNAFEFEPALADVKGKS LKDRRKKALERISPENIFKDKNNRIKEFAKGISAYSGANLTDGDFDGAKEELDHI IPRSHKKYGTLNDEANLICVTRGDNKNKGNRIFCLRDLADNYKLKQFETTDDLE IEKKIADTIWDANKKDFKFGNYRSFINLTPQEQKAFRHALFLADENPIKQAVIRA INNRNRTFVNGTQRYFAEVLANNIYLRAKKENLNTDKISFDYFGIPTIGNGRGIA EIRQLYEKVDSDIQAYAKGDKPQASYSHLIDAMLAFCIAADEHRNDGSIGLEID KNYSLYPLDKNTGEVFTKDIFSQIKITDNEFSDKKLVRKKAIEGFNTHRQMTRD GIYAENYLPILIHKELNEVRKGYTWKNSEEIKIFKGKKYDIQQLNNLVYCLKFV DKPISIDIQISTLEELRNILTTNNIAATAEYYYINLKTQKLHEYYIENYNTALGYK KYSKEMEFLRSLAYRSERVKIKSIDDVKQVLDKDSNFIIGKITLPFKKEWQRLYR EWQNTTIKDDYEFLKSFFNVKSITKLHKKVRKDFSLPISTNEGKFLVKRKTWDN NFIYQILNDSDSRADGTKPFIPAFDISKNEIVEAIIDSFTSKNIFWLPKNIELQKVD NKNIFAIDTSKWFEVETPSDLRDIGIATIQYKIDNNSRPKVRVKLDYVIDDDSKIN YFMNHSLLKSRYPDKVLEILKQSTIIEFESSGFNKTIKEMLGMKLAGIYNETSNN
Lactobacillus buchneri Cas9 SEQ ID NO: 16	MKVNNYHIGLDIGTSSIGWVAIGKDGKPLRVKGKTAIGARLFQEGNPAADRRM FRTTRRRLSRRKWRLKLLEEIFDPYITPVDSTFFARLKQSNLSPKDSRKEFKGSM LFPDLTDMQYHKNYPTIYHLRHALMTQDKKFDIRMVYLAIHHIVKYRGNFLNS TPVDSFKASKVDFVDQFKKLNELYAAINPEESFKINLANSEDIGHQFLDPSIRKF DKKKQIPKIVPVMMNDKVTDRLNGKIASEIIHAILGYKAKLDVVLQCTPVDSKP WALKFDDEDIDAKLEKILPEMDENQQSIVAILQNLYSQVTLNQIVPNGMSLSES MIEKYNDHHDHLKLYKKLIDQLADPKKKAVLKKAYSQYVGDDGKVIEQAEFW SSVKKNLDDSELSKQIMDLIDAEKFMPKQRTSQNGVIPHQLHQRELDEIIEHQSK YYPWLVEINPNKHDLHLAKYKIEQLVAFRVPYYVGPMITPKDQAESAETVFSW MERKGTETGQITPWNFDEKVDRKASANRFIKRMTTKDTYLIGEDVLPDESLLYE KFKVLNELNMVRVNGKLLKVADKQAIFQDLFENYKHVSVKKLQNYIKAKTGL PSDPEISGLSDPEHFNNSLGTYNDFKKLFGSKVDEPDLQDDFEKIVEWSTVFEDK KILREKLNEITWLSDQQKDVLESSRYQGWGRLSKKLLTGIVNDQGERIIDKLWN TNKNFMQIQSDDDFAKRIHEANADQMQAVDVEDVLADAYTSPQNKKAIRQVV KVVDDIQKAMGGVAPKYISIEFTRSEDRNPRRTISRQRQLENTLKDTAKSLAKSI NPELLSELDNAAKSKKGLTDRLYLYFTQLGKDIYTGEPINIDELNKYDIDHILPQ AFIKDNSLDNRVLVLTAVNNGKSDNVPLRMFGAKMGHFWKQLAEAGLISKRK LKNLQTDPDTISKYAMHGFIRRQLVETSQVIKLVANILGDKYRNDDTKIIEITAR MNHQMRDEFGFIKNREINDYHHAFDAYLTAFLGRYLYHRYIKLRPYFVYGDFK KFREDKVTMRNFNFLHDLTDDTQEKIADAETGEVIWDRENSIQQLKDVYHYKF MLISHEVYTLRGAMFNQTVYPASDAGKRKLIPVKADRPVNVYGGYSGSADAY MAIVRIHNKKGDKYRVVGVPMRALDRLDAAKNVSDADFDRALKDVLAPQLT KTKKSRKTGEITQVIEDFEIVLGKVMYRQLMIDGDKKFMLGSSTYQYNAKQLV LSDQSVKTLASKGRLDPLQESMDYNNVYTEILDKVNQYFSLYDMNKFRHKLN LGFSKFISFPNHNVLDGNTKVSSGKREILQEILNGLHANPTFGNLKDVGITTPFG QLQQPNGILLSDETKIRYQSPTGLFERTVSLKDL
Listeria innocua Cas9 SEQ ID NO: 17	MKKPYTIGLDIGTNSVGWAVLTDQYDLVKRKMKIAGDSEKKQIKKNFWGVRL FDEGQTAADRRMARTARRRIERRRNRISYLQGIFAEEMSKTDANFFCRLSDSFY VDNEKRNSRHPFFATIEEEVEYHKNYPTIYHLREELVNSSEKADLRLVYLALAHI IKYRGNFLIEGALDTQNTSVDGIYKQFIQTYNQVFASGIEDGSLKKLEDNKDVA KILVEKVTRKEKLERILKLYPGEKSAGMFAQFISLIVGSKGNFQKPFDLIEKSDIE CAKDSYEEDLESLLALIGDEYAELFVAAKNAYSAVVLSSIITVAETETNAKLSAS MIERFDTHEEDLGELKAFIKLHLPKHYEEIFSNTEKHGYAGYIDGKTKQADFYK YMKMTLENIEGADYFIAKIEKENFLRKQRTFDNGAIPHQLHLEELEAILHQQAK YYPFLKENYDKIKSLVTFRIPYFVGPLANGQSEFAWLTRKADGEIRPWNIEEKV DFGKSAVDFIEKMTNKDTYLPKENVLPKHSLCYQKYLVYNELTKVRYINDQGK TSYFSGQEKEQIFNDLFKQKRKVKKKDLELFLRNMSHVESPTIEGLEDSFNSSYS TYHDLLKVGIKQEILDNPVNTEMLENIVKILTVFEDKRMIKEQLQQFSDVLDGV

	VLKKLERRHYTGWGRLSAKLLMGIRDKQSHLTILDYLMNDDGLNRNLMQLIN DSNLSFKSIIEKEQVTTADKDIQSIVADLAGSPAIKKGILQSLKIVDELVSVMGYP PQTIVVEMARENQTTGKGKNNSRPRYKSLEKAIKEFGSQILKEHPTDNQELRNN RLYLYYLQNGKDMYTGQDLDIHNLSNYDIDHIVPQSFITDNSIDNLVLTSSAGN REKGDDVPPLEIVRKRKVFWEKLYQGNLMSKRKFDYLTKAERGGLTEADKAR FIHRQLVETRQITKNVANILHQRFNYEKDDHGNTMKQVRIVTLKSALVSQFRKQ FQLYKVRDVNDYHHAHDAYLNGVVANTLLKVYPQLEPEFVYGDYHQFDWFK ANKATAKKQFYTNIMLFFAQKDRIIDENGEILWDKKYLDTVKKVMSYRQMNIV KKTEIQKGEFSKATIKPKGNSSKLIPRKTNWDPMKYGGLDSPNMAYAVVIEYA KGKNKLVFEKKIIRVTIMERKAFEKDEKAFLEEQGYRQPKVLAKLPKYTLYECE EGRRRMLASANEAQKGNQQVLPNHLVTLLHHAANCEVSDGKSLDYIESNREM FAELLAHVSEFAKRYTLAEANLNKINQLFEQNKEGDIKAIAQSFVDLMAFNAM GAPASFKFFETTIERKRYNNLKELLNSTIIYQSITGLYESRKRLDD
L. pneumophilia Cas9 SEQ ID NO: 18	MESSQILSPIGIDLGGKFTGVCLSHLEAFAELPNHANTKYSVILIDHNNFQLSQA QRRATRHRVRNKKRNQFVKRVALQLFQHILSRDLNAKEETALCHYLNNRGYT YVDTDLDEYIKDETTINLLKELLPSESEHNFIDWFLQKMQSSEFRKILVSKVEEK KDDKELKNAVKNIKNFITGFEKNSVEGHRHRKVYFENIKSDITKDNQLDSIKKKI PSVCLSNLLGHLSNLQWKNLHRYLAKNPKQFDEQTFGNEFLRMLKNFRHLKGS QESLAVRNLIQQLEQSQDYISILEKTPPEITIPPYEARTNTGMEKDQSLLLNPEKL NNLYPNWRNLIPGIIDAHPFLEKDLEHTKLRDRKRIISPSKQDEKRDSYILQRYLD LNKKIDKFKIKKQLSFLGQGKQLPANLIETQKEMETHFNSSLVSVLIQIASAYNK EREDAAQGIWFDNAFSLCELSNINPPRKQKILPLLVGAILSEDFINNKDKWAKFK IFWNTHKIGRTSLKSKCKEIEEARKNSGNAFKIDYEEALNHPEHSNNKALIKIIQT IPDIIQAIQSHLGHNDSQALIYHNPFSLSQLYTILETKRDGFHKNCVAVTCENYW RSQKTEIDPEISYASRLPADSVRPFDGVLARMMQRLAYEIAMAKWEQIKHIPDN SSLLIPIYLEQNRFEFEESFKKIKGSSSDKTLEQAIEKQNIQWEEKFQRIINASMNI CPYKGASIGGQGEIDHIYPRSLSKKHFGVIFNSEVNLIYCSSQGNREKKEEHYLL EHLSPLYLKHQFGTDNVSDIKNFISQNVANIKKYISFHLLTPEQQKAARHALFLD YDDEAFKTITKFLMSQQKARVNGTQKFLGKQIMEFLSTLADSKQLQLEFSIKQIT AEEVHDHRELLSKQEPKLVKSRQQSFPSHAIDATLTMSIGLKEFPQFSQELDNS WFINHLMPDEVHLNPVRSKEKYNKPNISSTPLFKDSLYAERFIPVWVKGETFAIG FSEKDLFEIKPSNKEKLFTLLKTYSTKNPGESLQELQAKSKAKWLYFPINKTLAL EFLHHYFHKEIVTPDDTTVCHFINSLRYYTKKESITVKILKEPMPVLSVKFESSKK NVLGSFKHTIALPATKDWERLFNHPNFLALKANPAPNPKEFNEFIRKYFLSDNN PNSDIPNNGHNIKPQKHKAVRKVFSLPVIPGNAGTMMRIRRKDNKGQPLYQLQ TIDDTPSMGIQINEDRLVKQEVLMDAYKTRNLSTIDGINNSEGQAYATFDNWLT LPVSTFKPEIIKLEMKPHSKTRRYIRITQSLADFIKTIDEALMIKPSDSIDDPLNMP NEIVCKNKLFGNELKPRDGKMKIVSTGKIVTYEFESDSTPQWIQTLYVTQLKKQ P

N. lactamica Cas9 SEQ ID NO: 19	MAAFKPNPMNYILGLDIGIASVGWAMVEVDEEENPIRLIDLGVRVFERAEVPKT GDSLAMARRLARSVRRLTRRRAHRLLRARRLLKREGVLQDADFDENGLVKSL PNTPWQLRAAALDRKLTCLEWSAVLLHLVKHRGYLSQRKNEGETADKELGAL LKGVADNAHALQTGDFRTPAELALNKFEKESGHIRNQRGDYSHTFSRKDLQAE LNLLFEKQKEFGNPHVSDGLKEDIETLLMAQRPALSGDAVQKMLGHCTFEPAE PKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSKLTYA QARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKS PLNLSTELQDEIGTAFSLFKTDKDITGRLKDRVQPEILEALLKHISFDKFVQISLK ALRRIVPLMEQGKRYDEACAEIYGDHYCKKNAEEKIYLPPIPADEIRNPVVLRA LSQARKVINCVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENRKDREKAA AKFREYFPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLVRLNEKGYVEIDH ALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVE TSRFPRSKKQRILLQKFDEEGFKERNLNDTRYVNRFLCQFVADHILLTGKGKRR VFASNGQITNLLRGFWGLRKVRTENDRHHALDAVVVACSTVAMQQKITRFVR YKEMNAFDGKTIDKETGEVLHQKAHFPQPWEFFAQEVMIRVFGKPDGKPEFEE ADTPEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGQGHMETVKSAKR LDEGISVLRVPLTQLKLKGLEKMVNREREPKLYDALKAQLETHKDDPAKAFAE PFYKYDKAGSRTQQVKAVRIEQVQKTGVWVRNHNGIADNATMVRVDVFEKG GKYYLVPIYSWQVAKGILPDRAVVAFKDEEDWTVMDDSFEFRFVLYANDLIKL TAKKNEFLGYFVSLNRATGAIDIRTHDTDSTKGKNGIFQSVGVKTALSFQKNQI DELGKEIRPCRLKKRPPVR
N. meningitides Cas9 SEQ ID NO: 20	MAAFKPNPINYILGLDIGIASVGWAMVEIDEDENPICLIDLGVRVFERAEVPKTG DSLAMARRLARSVRRLTRRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPN TPWQLRAAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKG VADNAHALQTGDFRTPAELALNKFEKESGHIRNQRGDYSHTFSRKDLQAELILL FEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDAVQKMLGHCTFEPAEPKAA KNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSKLTYAQARK LLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNLS PELQDEIGTAFSLFKTDEDITGRLKDRIQPEILEALLKHISFDKFVQISLKALRRIV PLMEQGKRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQARK VINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENRKDREKAAAKFREY FPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLGRLNEKGYVEIDHALPFSRT WDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSRFPRS KKQRILLQKFDEDGFKERNLNDTRYVNRFLCQFVADRMRLTGKGKKRVFASN GQITNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRYKEMN AFDGKTIDKETGEVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTPEK LRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGQGHMETVKSAKRLDEGVS VLRVPLTQLKLKDLEKMVNREREPKLYEALKARLEAHKDDPAKAFAEPFYKY DKAGNRTQQVKAVRVEQVQKTGVWVRNHNGIADNATMVRVDVFEKGDKYY LVPIYSWQVAKGILPDRAVVQGKDEEDWQLIDDSFNFKFSLHPNDLVEVITKKA RMFGYFASCHRGTGNINIRIHDLDHKIGKNGILEGIGVKTALSFQKYQIDELGKEI RPCRLKKRPPVR
B. longum Cas9 SEQ ID NO: 21	MLSRQLLGASHLARPVSYSYNVQDNDVHCSYGERCFMRGKRYRIGIDVGLNSV GLAAVEVSDENSPVRLLNAQSVIHDGGVDPQKNKEAITRKNMSGVARRTRRM RRRKRERLHKLDMLLGKFGYPVIEPESLDKPFEEWHVRAELATRYIEDDELRRE SISIALRHMARHRGWRNPYRQVDSLISDNPYSKQYGELKEKAKAYNDDATAAE EESTPAQLVVAMLDAGYAEAPRLRWRTGSKKPDAEGYLPVRLMQEDNANELK QIFRVQRVPADEWKPLFRSVFYAVSPKGSAEQRVGQDPLAPEQARALKASLAF QEYRIANVITNLRIKDASAELRKLTVDEKQSIYDQLVSPSSEDITWSDLCDFLGF KRSQLKGVGSLTEDGEERISSRPPRLTSVQRIYESDNKIRKPLVAWWKSASDNE HEAMIRLLSNTVDIDKVREDVAYASAIEFIDGLDDDALTKLDSVDLPSGRAAYS VETLQKLTRQMLTTDDDLHEARKTLFNVTDSWRPPADPIGEPLGNPSVDRVLK NVNRYLMNCQQRWGNPVSVNIEHVRSSFSSVAFARKDKREYEKNNEKRSIFRS SLSEQLRADEQMEKVRESDLRRLEAIQRQNGQCLYCGRTITFRTCEMDHIVPRK

	GVGSTNTRTNFAAVCAECNRMKSNTPFAIWARSEDAQTRGVSLAEAKKRVTM FTFNPKSYAPREVKAFKQAVIARLQQTEDDAAIDNRSIESVAWMADELHRRID WYFNAKQYVNSASIDDAEAETMKTTVSVFQGRVTASARRAAGIEGKIHFIGQQ SKTRLDRRHHAVDASVIAMMNTAAAQTLMERESLRESQRLIGLMPGERSWKE YPYEGTSRYESFHLWLDNMDVLLELLNDALDNDRIAVMQSQRYVLGNSIAHD ATIHPLEKVPLGSAMSADLIRRASTPALWCALTRLPDYDEKEGLPEDSHREIRV HDTRYSADDEMGFFASQAAQIAVQEGSADIGSAIHHARVYRCWKTNAKGVRK YFYGMIRVFQTDLLRACHDDLFTVPLPPQSISMRYGEPRVVQALQSGNAQYLG SLVVGDEIEMDFSSLDVDGQIGEYLQFFSQFSGGNLAWKHWVVDGFFNQTQLR IRPRYLAAEGLAKAFSDDVVPDGVQKIVTKQGWLPPVNTASKTAVRIVRRNAF GEPRLSSAHHMPCSWQWRHE
A. muciniphila Cas9 SEQ ID NO: 22	MSRSLTFSFDIGYASIGWAVIASASHDDADPSVCGCGTVLFPKDDCQAFKRREY RRLRRNIRSRRVRIERIGRLLVQAQIITPEMKETSGHPAPFYLASEALKGHRTLAP IELWHVLRWYAHNRGYDNNASWSNSLSEDGGNGEDTERVKHAQDLMDKHGT ATMAETICRELKLEEGKADAPMEVSTPAYKNLNTAFPRLIVEKEVRRILELSAPL IPGLTAEIIELIAQHHPLTTEQRGVLLQHGIKLARRYRGSLLFGQLIPRFDNRIISR CPVTWAQVYEAELKKGNSEQSARERAEKLSKVPTANCPEFYEYRMARILCNIR ADGEPLSAEIRRELMNQARQEGKLTKASLEKAISSRLGKETETNVSNYFTLHPD SEEALYLNPAVEVLQRSGIGQILSPSVYRIAANRLRRGKSVTPNYLLNLLKSRGE SGEALEKKIEKESKKKEADYADTPLKPKYATGRAPYARTVLKKVVEEILDGEDP TRPARGEAHPDGELKAHDGCLYCLLDTDSSVNQHQKERRLDTMTNNHLVRHR MLILDRLLKDLIQDFADGQKDRISRVCVEVGKELTTFSAMDSKKIQRELTLRQK SHTDAVNRLKRKLPGKALSANLIRKCRIAMDMNWTCPFTGATYGDHELENLEL EHIVPHSFRQSNALSSLVLTWPGVNRMKGQRTGYDFVEQEQENPVPDKPNLHI CSLNNYRELVEKLDDKKGHEDDRRRKKKRKALLMVRGLSHKHQSQNHEAMK EIGMTEGMMTQSSHLMKLACKSIKTSLPDAHIDMIPGAVTAEVRKAWDVFGVF KELCPEAADPDSGKILKENLRSLTHLHHALDACVLGLIPYIIPAHHNGLLRRVLA MRRIPEKLIPQVRPVANQRHYVLNDDGRMMLRDLSASLKENIREQLMEQRVIQ HVPADMGGALLKETMQRVLSVDGSGEDAMVSLSKKKDGKKEKNQVKASKLV GVFPEGPSKLKALKAAIEIDGNYGVALDPKPVVIRHIKVFKRIMALKEQNGGKP VRILKKGMLIHLTSSKDPKHAGVWRIESIQDSKGGVKLDLQRAHCAVPKNKTH ECNWREVDLISLLKKYQMKRYPTSYTGTPR
O. laneus Cas9 SEQ ID NO: 23	METTLGIDLGTNSIGLALVDQEEHQILYSGVRIFPEGINKDTIGLGEKEESRNATR RAKRQMRRQYFRKKLRKAKLLELLIAYDMCPLKPEDVRRWKNWDKQQKSTV RQFPDTPAFREWLKQNPYELRKQAVTEDVTRPELGRILYQMIQRRGFLSSRKGK EEGKIFTGKDRMVGIDETRKNLQKQTLGAYLYDIAPKNGEKYRFRTERVRARY TLRDMYIREFEIIWQRQAGHLGLAHEQATRKKNIFLEGSATNVRNSKLITHLQA KYGRGHVLIEDTRITVTFQLPLKEVLGGKIEIEEEQLKFKSNESVLFWQRPLRSQ KSLLSKCVFEGRNFYDPVHQKWIIAGPTPAPLSHPEFEEFRAYQFINNIIYGKNEH LTAIQREAVFELMCTESKDFNFEKIPKHLKLFEKFNFDDTTKVPACTTISQLRKL FPHPVWEEKREEIWHCFYFYDDNTLLFEKLQKDYALQTNDLEKIKKIRLSESYG NVSLKAIRRINPYLKKGYAYSTAVLLGGIRNSFGKRFEYFKEYEPEIEKAVCRIL KEKNAEGEVIRKIKDYLVHNRFGFAKNDRAFQKLYHHSQAITTQAQKERLPET GNLRNPIVQQGLNELRRTVNKLLATCREKYGPSFKFDHIHVEMGRELRSSKTER EKQSRQIRENEKKNEAAKVKLAEYGLKAYRDNIQKYLLYKEIEEKGGTVCCPY TGKTLNISHTLGSDNSVQIEHIIPYSISLDDSLANKTLCDATFNREKGELTPYDFY QKDPSPEKWGASSWEEIEDRAFRLLPYAKAQRFIRRKPQESNEFISRQLNDTRYI SKKAVEYLSAICSDVKAFPGQLTAELRHLWGLNNILQSAPDITFPLPVSATENHR EYYVITNEQNEVIRLFPKQGETPRTEKGELLLTGEVERKVFRCKGMQEFQTDVS DGKYWRRIKLSSSVTWSPLFAPKPISADGQIVLKGRIEKGVFVCNQLKQKLKTG LPDGSYWISLPVISQTFKEGESVNNSKLTSQQVQLFGRVREGIFRCHNYQCPASG ADGNFWCTLDTDTAQPAFTPIKNAPPGVGGGQIILTGDVDDKGIFHADDDLHYE LPASLPKGKYYGIFTVESCDPTLIPIELSAPKTSKGENLIEGNIWVDEHTGEVRFD PKKNREDQRHHAIDAIVIALSSQSLFQRLSTYNARRENKKRGLDSTEHFPSPWP

	GFAQDVRQSVVPLLVSYKQNPKTLCKISKTLYKDGKKIHSCGNAVRGQLHKET VYGQRTAPGATEKSYHIRKDIRELKTSKHIGKVVDITIRQMLLKHLQENYHIDIT QEFNIPSNAFFKEGVYRIFLPNKHGEPVPIKKIRMKEELGNAERLKDNINQYVNP RNNHHVMIYQDADGNLKEEIVSFWSVIERQNQGQPIYQLPREGRNIVSILQINDT FLIGLKEEEPEVYRNDLSTLSKHLYRVQKLSGMYYTFRHHLASTLNNEREEFRI QSLEAWKRANPVKVQIDEIGRITFLNGPLC

[0105] In some embodiments, a nucleic acid sequence encoding a dCas endonuclease is a codon optimized dCas. An example of a codon optimized sequence, is in this instance, a sequence optimized for expression in, without limitation, a eukaryote, animal, and/or mammal e.g., a human (i.e. being optimized for expression in humans); see, e.g., *Sa*Cas9 human codon optimized sequence in WO 2014/093622, incorporated by reference herein in its entirety.

[0106] In some embodiments, a dCas endonuclease for use in the system provided herein is a variant Cas endonuclease comprising mutations which cause the endonuclease to lack cleavage activity or substantially lack cleavage activity as compared to its corresponding wild type Cas endonuclease. For example, with reference to WO 2017/091630, incorporated herein by reference in its entirety, in one embodiment disclosed herein, the Cas9 active sites (10 and 840) can be mutated to Alanine (D10A and H840A) to eliminate the cleavage activity of *Streptococcus pyogenes* Cas9, producing nuclease-deficient or dead Cas9 (i.e., dCas9). The RuvC domain is distributed among 3 non-contiguous portions of the dCas9 primary structure (residues 1-60, 719-775, and 910-1099). The Rec lobe is composed of residues 61-718. The HNH domain is composed of residues 776-909. The PAM-ID domain is composed of residues 1100-1368. The REC lobe can be considered the structural scaffold for recognition of the sgRNA and target DNA/RNA. The NUC lobe contains the two nuclease domains (HNH and RuvC), plus the PAM-interaction domain (PAM-ID), which recognizes an optional PAM sequence. In this prior work, for example and without limitation, an about 98-nucleotide sgRNA, is typically divided into two major structural components: the first

contains the target-specific guide or "spacer" segment (nucleotides 1-20) plus the repeat-tetraloop-anti-repeat and stem-loop 1 (SL1) regions; the second contains stem-loops 2 and 3 (SL2, SL3). Accordingly, the guide-through-SL1 RNA segment is bound mainly by the Cas9 REC lobe and the SL2-SL3 segment is bound mainly by the NUC lobe.

[0107] In some embodiments of the dCas9 used in the system disclosed herein, a minimal (i.e., with as few nucleotide base pairs as possible) construct of Cas9 is engineered that will recognize a target RNA sequence with high affinity. In some embodiments, the smallest construct encoding dCas9 will be a REC-only construct. In some embodiments, the constructs will comprise less minimized constructs lacking the HNH, PAM-ID, parts of each domain, lacking both of each domain, or combinations thereof. In some embodiments, the HNH domain will be excised by inserting a five-residue flexible linker between residues 775 and 909 (ΔHNH). In some embodiments, all or part of the PAM-ID are removed. In some embodiments, truncating Cas9 at residue 1098 (ΔPAM-ID #1), fusing residues 1138 and 1345 with an 8-residue linker (ΔPAM-ID #2), or fusing residues 1138 with 1200 and 1218 with 1339 (with 5-residue and 2-residue linkers, respectively: ΔPAM-ID #3) are used to remove all or part of the PAM-ID. The ΔPAM-ID #2 and 3 constructs will retain elements of the PAM-ID that contribute to binding of the sgRNA repeat-anti-repeat (residues 1099-1138) and SL2-SL3 (residues 1200-1218 and 1339-1368) segments. In some embodiments, the HNH deletion will be combined with the three PAM-ID deletions. In some embodiments, Cas9 variants which lack or substantially lack nuclease and/or cleavage activity according to WO 2016/19655, incorporated herein by reference in its entirety, are examples of dCas9 used in the recombinant expression systems disclosed herein.

[0108] Accordingly for use in the recombinant expression systems disclosed herein are nucleic acid sequences encoding dCas – ADAR deaminase domain fusion proteins. In one embodiment, dCas9 is fused to a catalytically active ADAR deaminase domain. In the context of such systems a corresponding extended single guide RNA (esgRNA) is used to target and edit adenosines of the target RNA. The system generates recombinant proteins with effector deaminase enzymes capable of performing ribonucleotide base modification to alter how sequence of the RNA molecule is recognized by cellular machinery. In one embodiment the dCas and the ADAR deaminase domain are separated by a linker. In another

embodiment, the linker is, without limitation, an XTEN linker which is a flexible linker used to isolate adjacent proteins domains. XTEN linkers are known in the art and can be found for example in WO 2013/130684, incorporated herein by reference in its entirety herein.

[0109] RNA editing is a natural process whereby the diversity of gene products of a given sequence is increased by minor modification in the RNA. Typically, the modification involves the conversion of adenosine (A) to inosine (I), resulting in an RNA sequence which is different from that encoded by the genome. RNA modification is generally ensured by the ADAR enzyme, whereby the pre-RNA target forms an imperfect duplex RNA by base-pairing between the exon that contains the adenosine to be edited and an intronic non-coding element. A classic example of A-I editing is the glutamate receptor GluR-B mRNA, whereby the change results in modified conductance properties of the channel (Higuchi M, et al. Cell. 1993;75: 1361-70).

[0110] For the purposes of the present disclosure, ADAR (Adenosine deaminase acting on RNA) deaminase domains can be ADAR 1, ADAR 2, or ADAR 3 deaminase domains. See Nishikura, K. A-to-I editing of coding and non-coding RNAs by ADARs. Nat Rev Mol Cell Biol 17, 83-96, doi:10.1038/nrm.2015.4 (2016).

[0111] In some embodiments, the ADAR deaminase domain is derived from all or part of ADAR1 (Uniprot P55265). A non-limiting exemplary sequence of ADAR1 is provided below (SEQ ID NO: 24):

MAEIKEKICDYLFNVSDSSALNLAKNIGLTKARDINAVLIDMERQGDVYRQGTTPPI
WHLTDKKRERMQIKRNTNSVPETAPAAIPETKRNAEFLTCNIPTSNASNNMVTTEKV
ENGQEPVIKLENRQEARPEPARLKPPVHYNGPSKAGYVDFENGQWATDDIPDDLNSI
RAAPGEFRAIMEMPSFYSHGLPRCSPYKKLTECQLKNPISGLLEYAQFASQTCEFNMI
EQSGPPHEPRFKFQVVINGREFPPAEAGSKKVAKQDAAMKAMTILLEEAKAKDSGK
SEESSHYSTEKESEKTAESQTPTPSATSFFSGKSPVTTLLECMHKLGNSCEFRLLSKEG
PAHEPKFQYCVAVGAQTFPSVSAPSKKVAKQMAAEEAMKALHGEATNSMASDNQP
EGMISESLDNLESMMPNKVRKIGELVRYLNTNPVGGLLEYARSHGFAAEFKLVDQS
GPPHEPKFVYQAKVGGRWFPAVCAHSKKQGKQEAADAALRVLIGENEKAERMGFT
EVTPVTGASLRRTMLLLSRSPEAQPKTLPLTGSTFHDQIAMLSHRCFNTLTNSFQPSLL

GRKILAAIIMKKDSEDMGVVVSLGTGNRCVKGDSLSLKGETVNDCHAEIISRRGFIRF
LYSELMKYNSQTAKDSIFEPAKGGEKLQIKKTVSFHLYISTAPCGDGALFDKSCSDRA
MESTESRHYPVFENPKQGKLRTKVENGEGTIPVESSDIVPTWDGIRLGERLRTMSCSD
KILRWNVLGLQGALLTHFLQPIYLKSVTLGYLFSQGHLTRAICCRVTRDGSAFEDGLR
HPFIVNHPKVGRVSIYDSKRQSGKTKETSVNWCLADGYDLEILDGTRGTVDGPRNEL
SRVSKKNIFLLFKKLCSFRYRRDLLRLSYGEAKKAARDYETAKNYFKKGLKDMGYG
NWISKPQEEKNFYLCPV

[0112] In some embodiments, the ADAR deaminase domain is derived from all or part of ADAR2 (Uniprot P78563). A non-limiting exemplary sequence of ADAR2 is provided below (SEQ ID NO: 25):

MDIEDEENMSSSSTDVKENRNLDNVSPKDGSTPGPGEGSQLSNGGGGGPGRKRPLEE
GSNGHSKYRLKKRRKTPGPVLPKNALMQLNEIKPGLQYTLLSQTGPVHAPLFVMSV
EVNGQVFEGSGPTKKKAKLHAAEKALRSFVQFPNASEAHLAMGRTLSVNTDFTSDQ
ADFPDTLFNGFETPDKAEPPFYVGSNGDDSFSSSGDLSLSASPVPASLAQPPLPVLPPF
PPPSGKNPVMILNELRPGLKYDFLSESGESHAKSFVMSVVVDGQFFEGSGRNKKLAK
ARAAQSALAAIFNLHLDQTPSRQPIPSEGLQLHLPQVLADAVSRLVLGKFGDLTDNFS
SPHARRKVLAGVVMTTGTDVKDAKVISVSTGTKCINGEYMSDRGLALNDCHAEIISR
RSLLRFLYTQLELYLNNKDDQKRSIFQKSERGGFRLKENVQFHLYISTSPCGDARIFSP
HEPILEEPADRHPNRKARGQLRTKIESGEGTIPVRSNASIQTWDGVLQGERLLTMSCS
DKIARWNVVGIQGSLLSIFVEPIYFSSIILGSLYHGDHLSRAMYQRISNIEDLPPLYTLN
KPLLSGISNAEARQPGKAPNFSVNWTVGDSAIEVINATTGKDELGRASRLCKHALYC
RWMRVHGKVPSHLLRSKITKPNVYHESKLAAKEYQAAKARLFTAFIKAGLGAWVE
KPTEQDQFSLTP

[0113] In some embodiments, the ADAR deaminase domain is derived from all or part of ADAR3 (Uniprot Q9NS39): A non-limiting exemplary sequence of ADAR2 is provided below (SEQ ID NO: 26):

MASVLGSGRGSGGLSSQLKCKSKRRRRRRSKRKDKVSILSTFLAPFKHLSPGITNTED
DDTLSTSSAEVKENRNVGNLAARPPPSGDRARGGAPGAKRKRPLEEGNGGHLCKLQ
LVWKKLSWSVAPKNALVQLHELRPGLQYRTVSQTGPVHAPVFAVAVEVNGLTFEG

TGPTKKKAKMRAAELALRSFVQFPNACQAHLAMGGGPGPGTDFTSDQADFPDTLFQ
EFEPPAPRPGLAGGRPGDAALLSAAYGRRRLLCRALDLVGPTPATPAAPGERNPVVL
LNRLRAGLRYVCLAEPAERRARSFVMAVSVDGRTFEGSGRSKKLARGQAAQAALQ
ELFDIQMPGHAPGRARRTPMPQEFADSISQLVTQKFREVTTDLTPMHARHKALAGIV
MTKGLDARQAQVVALSSGTKCISGEHLSDQGLVVNDCHAEVVARRAFLHFLYTQLE
LHLSKRREDSERSIFVRLKEGGYRLRENILFHLYVSTSPCGDARLHSPYEITTDLHSSK
HLVRKFRGHLRTKIESGEGTVPVRGPSAVQTWDGVLLGEQLITMSCTDKIARWNVL
GLQGALLSHFVEPVYLQSIVVGSLHHTGHLARVMSHRMEGVGQLPASYRHNRPLLS
GVSDAEARQPGKSPPFSMNWVVGSADLEIINATTGRRSCGGPSRLCKHVLSARWAR
LYGRLSTRTPSPGDTPSMYCEAKLGAHTYQSVKQQLFKAFQKAGLGTWVRKPPEQQ
QFLLTL

[0114] In some embodiments, ADAR domains can include mutations which result in increased catalytic activity compared to wild type ADAR domains. In some embodiments, the catalytically active deaminase domain (DD) is derived from a wildtype human ADAR2 or a human ADAR2 DD bearing a mutation (E488Q) that increases enzymatic activity and affinity for RNA substrate (Phelps et al., Jan 2015, Nuc. Acid Res., 43(2): 1123-1132; Kuttan & Bass, Nov 2012, PNAS 109(48): E3295-E3304).

[0115] Because the catalytic domain of ADAR2, independent of its RNA recognition motif, preferably deaminates unpaired adenosine residues in dsRNA regions, Applicants modified the structure of the single guide RNA (sgRNA) component of the system disclosed herein to improve substrate specificity to single-nucleotide resolution. It has been reported that gRNAs engineered with supplementary 3' terminal cassettes maintain their targeting capacity in live cells (Konermann et al. Jan 2015, Nature, 517: 583-588).

[0116] Applicants developed a CRISPR/Cas-mediated RNA editing (CREDIT) platform based on the strategic modification of the system's sgRNA structure comprising an additional region of homology capable of base pairing with target RNA over the desired site of editing. Such a modification to the sgRNA structure generates the disclosed system's extended sgRNA (i.e., esgRNA), and results in an A-to- C mismatch with a target transcript generating a 'pseudo-dsRNA' substrate to be edited at the bulged adenosine (see FIG. 1A). The CREDIT platform and the systems disclosed herein thus provides the ability to target virtually any

adenosine in the transcriptome to direct conversion to inosine (i.e., A – I RNA editing), which is ultimately read by translational and splicing machinery as guanosine.

[0117] Due to its overall design simplicity as well as its fully encodable nature, the recombinant expression systems disclosed herein provide high utility and engineering versatility when compared to other similar RNA modifying systems and methods. Because dCas9 binds with picomolar affinity to the sgRNA scaffold sequence, and because this improved system uses dual guide architecture as per the extended single guide RNA i.e., esgRNA, structure, to increase both target affinity and specificity, direct RNA editing with minimal potential off-target editing events is efficiently achieved. In some embodiments, the esgRNA can be designed with a i) scaffold sequence and ii) a short extension sequence but without a spacer sequence.

[0118] In one embodiment, the esgRNA is composed of at least two regions, i) a region of homology capable of near-perfect RNA-RNA base pairing (i.e., a short extension sequence of homology to the target RNA) and ii) a dCas9-binding region (i.e., scaffold sequence). In one embodiment, the short extension sequence comprises a mismatch which forms an A-C mismatch with a target transcriptome and generates a 'pseudo-RNA' substrate to be edited at the bulged adenosine residue. As such, the homology region of the short extension sequence determines the specificity of the recombinant expression system disclosed herein, and in particular it determines specifically which RNA base in the cellular transcriptome is edited. The RNA base that is edited is distinguished by a mismatched adenosine residue among the homology region and the target RNA duplex. See FIG. 1A. The orientation of the homology region of the short extension sequence and the scaffold is flexible. In one embodiment, the scaffold sequence is located at the 5' end of the esgRNA. In another embodiment, the short extension sequence carrying the homology region capable of near-perfect RNA-RNA base pairing is located at the 3' end of the esgRNA. In another embodiment, the short extension sequence is located at the 5' end of the esgRNA. For the purposes of the present disclosure, the "3' end" or "5' end" refers in either scenario of the esgRNA to an end terminus of the esgRNA. In another embodiment, the esgRNA additionally comprises a third region, iii) a spacer sequence which comprises a second homology region to the target RNA. In one embodiment, the spacer sequence is located at the 5' end of the scaffold sequence. The

spacer sequence is complementary to the target RNA but does not require a mismatch to effect the A-I editing of the target RNA. In one embodiment, the spacer sequence is located on the 5' end of the scaffold sequence. In another embodiment, the short extension sequence is located on the 3' end of the scaffold sequence or on the 5' end of the spacer sequence. In another embodiment, the short extension sequence is located on an end terminus of the esgRNA. In another embodiment, the short extension sequence is continuous to the spacer sequence. In another embodiment, the short extension sequence is discontinuous to the spacer sequence. In another embodiment, the esgRNA comprising i-iii) in a 3' to 5' orientation.

[0119] In some embodiments, nucleoprotein complexes are complexed with a single guide RNA (sgRNA) or as disclosed herein an extended single guide RNA (esgRNA). In some embodiments, the single guide RNA or esgRNA carries extensions (other than and in addition to the short extension sequence of homology in the esgRNA capable of editing target adenosines) of secondary structures in the single guide RNA or esgRNA scaffold sequence. In some embodiments, the single guide RNA or esgRNA comprises one or more point mutations that improve expression levels of the single guide RNAs (or esgRNAs) via removal of partial or full transcription termination sequences or sequences that destabilize single guide RNAs (or esgRNAs) after transcription via action of trans-acting nucleases. In some embodiments, the single guide RNA (or esgRNA) comprises an alteration at the 5′ end which stabilizes said single guide RNA or esgRNA against degradation. In some embodiments, the single guide RNA or esgRNA comprises an alteration at the 5′ end which improves RNA targeting. In some embodiments, the alteration at the 5′ end of said single guide RNA or esgRNA is selected from the group consisting of 2′O-methyl, phosphorothioates, and thiophosphonoacetate linkages and bases. In some embodiments, the single guide RNA or esgRNA comprises 2′-fluorine, 2′O-methyl, and/or 2′-methoxyethyl base modifications in the spacer or scaffold region of the sgRNA or esgRNA to improve target recognition or reduce nuclease activity on the single guide RNA or esgRNA. In some embodiments, the single guide RNA comprises one or more methylphosphonate, thiophosponoaceteate, or phosphorothioate linkages that reduce nuclease activity on the target RNA.

[0120] In some embodiments, the single guide RNA or esgRNA can recognize the target RNA, for example, by hybridizing to the target RNA. In some embodiments, the single guide RNA or esgRNA comprises a sequence that is complementary to the target RNA. In some embodiments, the single guide RNA or esgRNA has a length that is, is about, is less than, or is more than, 10 nt, 20 nt, 30 nt, 40 nt, 50 nt, 60 nt, 70 nt, 80 nt, 90 nt, 100 nt, 110 nt, 120 nt, 130 nt, 140 nt, 150 nt, 160 nt, 170 nt, 180 nt, 190 nt, 200 nt, 300 nt, 400 nt, 500 nt, 1,000 nt, 2,000 nt, or a range between any two of the above values. In some embodiments, the single guide RNA or esgRNA can comprise one or more modified nucleotides.

[0121] In additional embodiments, a variety of RNA targets can be recognized by the single guide RNA or esgRNA. For example, a target RNA can be messenger RNA (mRNA), ribosomal RNA (rRNA), signal recognition particle RNA (SRP RNA), transfer RNA (tRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), antisense RNA (aRNA), long noncoding RNA (lncRNA), microRNA (miRNA), piwi-interacting RNA (piRNA), small interfering RNA (siRNA), short hairpin RNA (shRNA), retrotransposon RNA, viral genome RNA, viral noncoding RNA, or the like. In some embodiments, a target RNA can be an RNA involved in pathogenesis or a therapeutic target for conditions such as cancers, neurodegeneration, cutaneous conditions, endocrine conditions, intestinal diseases, infectious conditions, neurological disorders, liver diseases, heart disorders, autoimmune diseases, or the like.

[0122] In further embodiments, exemplary G to A mutation target RNA and corresponding diseases, conditions and/or syndromes to be treated are, without limitation:

[0123] SDHB (Succinate Dehydrogenase Complex Iron Sulfure Subunit B) for treating Paraganglioma, gastric stromal sarcoma, Paragangliomas 4, Pheochromocytoma, Paragangliomas 1, and/or Hereditary cancer-predisposing syndrome;

[0124] DPYD (Dihydropyrimidine Dehydrogenase) for treating Dihydropyrimidine dehydrogenase deficiency, Hirschsprung disease 1, Fluorouracil response, Pyrimidine analogues response - Toxicity/ADR, capecitabine response - Toxicity/ADR, fluorouracil response - Toxicity/ADR, and/or tegafur response - Toxicity/ADR;

[0125] MSH2 (mutS Homolog 2) for treating Lynch syndrome, tumor predisposition syndrome, and/or Turcot syndrome;

[0126] MSH6 (mutS Homolog 6) for treating Lynch syndrome;

[0127] DYSF (Dysferlin) for treating Miyoshi muscular dystrophy 1, and/or Limb-girdle muscular dystrophy -type 2B;

[0128] SCN1A (Sodium Voltage-Gated Channel Alpha Subunit 1) for treating Severe myoclonic epilepsy in infancy;

[0129] TTN (Titin) / TTN-AS1 for treating Primary dilated cardiomyopathy;

[0130] VHL (von Hippel-Lindau Tumor Suppressor) for treating Von Hippel-Lindau syndrome; and/or Hereditary cancer-predisposing syndrome;

[0131] MLH1 (mutL homolog 1) for treating Lynch syndrome, Hereditary cancer-predisposing syndrome, and/or tumor predisposition syndrome;

[0132] PDE6B (Phosphodiesterase 6B) for treating Retinitis pigmentosa and/or Retinitis pigmentosa 40;

[0133] CC2D2A (Coiled-coil and C2 Domain Containing 2A) for treating Familial aplasia of the vermis and/or Joubert syndrome 9;

[0134] FRAS1 (Fraser extracellular matrix complex subunit 1) for treating Cryptophthalmos syndrome;

[0135] DSP (Desmoplakin) for treating Arrhythmogenic right ventricular cardiomyopathy - type 8 and/or Cardiomyopathy;

[0136] PMS2 (PMS1 homolog 2, mismatch repair system component) for treating Lynch syndrome and/or tumor predisposition syndrome;

[0137] ASL (Argininosuccinate lyase) for treating Argininosuccinic aciduria;

[0138] ELN (Elastin) for treating Supravalvar aortic stenosis;

[0139] SLC26A4 (Solute Carrier Family 26 Member 4) for treating Enlarged vestibular aqueduct syndrome and/or Pendred's syndrome;

[0140] CFTR (Cystic Fibrosis Transmembrane Conductance Regulator) for treating Cystic Fibrosis;

[0141] CNGB3 (Cyclic Nucleotide Gated Channel Beta 3) for treating Achromatopsia 3;

[0142] FANCC (Fanconi Anemia Complementation Group C) – C9orf3 for treating Fanconi anemia and/or Hereditary cancer-predisposing syndrome;

[0143] PTEN (Phosphatase and Tensin homolog) for treating Hereditary cancer-predisposing syndrome, Bannayan-Riley-Ruvalcaba syndrome, Cowden syndrome, Breast cancer, Autism spectrum disorder, Head and neck squamous cell carcinoma, lung cancer, and/or prostate cancer;

[0144] ANO5 (Anoctamin 5) for treating Limb-girdle muscular dystrophy - type 2L, Gnathodiaphyseal dysplasmia, Miyoshi myopathy, and/or Miyoshi muscular dystrophy 3;

[0145] MYBPC3 (Myosin Binding Protein C, Cardiac) for treating Primary familial hypertrophic cardiomyopathy;

[0146] MEN1 (Menin 1) for treating Familial isolated hyperparathyroidism, multiple endocrine neoplasia, primary macronodular adrenal hyperplasia, and/or tumors;

[0147] ATM (ATM serine/threonine kinase) and/or ATM-C11orf65 for treating Ataxia-telangiectasia syndrome, and/or Hereditary cancer-predisposing syndrome;

[0148] PKP2 (Plakophilin 2) for treating Arrhythmogenic right ventricular cardiomyopathy - type 9 and/or Arrhythmogenic right ventricular cardiomyopathy;

[0149] PAH (Phenylalanine Hydroxylase) for treating Phenylketonuria;

[0150] GJB2 (Gap Junction Protein Beta 2) for treating Deafness, autosomal recessive 1A, Non-syndromic genetic deafness and/or Hearing impairment;

[0151] B3GLCT (beta 3-glucosyltransferase) for treating Peters plus syndrome;

[0152] BRCA2 (BRCA2, DNA repair associated) for treating Familial cancer of breast, Breast-ovarian cancer - familial 2, Hereditary cancer-predisposing syndrome, Fanconi anemia, complementation group D1, Hereditary breast and ovarian cancer syndrome, Hereditary cancer-predisposing syndrome, Breast-ovarian cancer - familial 1, and/or Hereditary breast and ovarian cancer syndrome;

[0153] MYH7 (Myosin Heavy Chain 7) for treating Primary dilated cardiomyopathy, Cardiomyopathy, and/or Cardiomyopathy - left ventricular noncompaction;

[0154] FBN1 (Fibrillin 1) for treating Marfan syndrome;

[0155] HEXA (Hexosaminidase Subunit Alpha) for treating Tay-Sachs disease;

[0156] TSC2 (TSC Complex Subunit 2) for treating Tuberous sclerosis 2, and/or Tuberous sclerosis syndrome;

[0157] CREBBP (CREB binding protein) for treating Rubinstein-Taybi syndrome;

[0158] CDH1 (Cadherin 1) for treating Hereditary diffuse gastric cancer, Tumor predisposition syndrome, and/or Hereditary cancer-predisposing syndrome;

[0159] SPG7 (SPG7, paraplegin matrix AAA peptidase subunit) for treating Spastic paraplegia 7;

[0160] BRCA1 (BRCA1, DNA repair associated) for treating Breast-ovarian cancer - familial 1, Hereditary breast and ovarian cancer syndrome, and/or Hereditary cancer-predisposing syndrome;

[0161] BRIP1 (BRCA1 Interacting Protein C-Terminal Helicase 1) for treating Familial cancer of breast and/or Tumor predisposition syndrome;

[0162] LDLR (Low Density Lipoprotein Receptor) and/or LDLR – MIR6886 for treating Familial hypercholesterolemia and/or Hypercholesterolaemia;

[0163] BCKDHA (Branced Chain Keto acid dehydrogenase E1, alpha polypeptide) for treating Maple syrup urine disease;

[0164] CHEK2 (Checkpoint Kinase 2) for treating Familial cancer of breast, Breast and colorectal cancer - susceptibility to, and/or Hereditary cancer-predisposing syndrome;

[0165] DMD (Dystrophin) for treating Becker muscular dystrophy, Duchenne muscular dystrophy, and/or Dilated cardiomyopathy 3B; and/or

[0166] IDUA (Iduronidase, alpha-L) for treating Hurler syndrome, Dysostosis multiplex, Mucopolysaccharidosis, MPS-I-H/S, and/or Mucopolysaccharidosis type I.

[0167] In some embodiments, the esgRNA comprises a short extension sequence of homology to the target RNA which is about 10-100 nucleotides in length, or about 10, 15-60, 20-50, or 25-40, or any range therebetween nucleotides in length. In some embodiments, the short extension sequence of the esgRNA, without limitation, comprising about 1 mismatch or 2, 3, 4, or 5 mismatches.

[0168] In some embodiments, the single guide RNA or esgRNA includes, but is not limited to including, sequences which bind or hybridize to target RNA, such as spacer sequences comprising additional regions of homology (in addition to the short extension sequence of homology disclosed herein) to the target RNA such that RNA recognition is supported with specificity and provides uniquely flexible and accessible manipulation of the genome. See WO 2017/091630 incorporated by reference in its entirety herein.

[0169] Non-limiting exemplary spacer sequences and extension sequences designed for esgRNA targeting the CFTR mRNA (cystic fibrosis transmembrane conductance regulator, Ref Seq: NM_000492) and the IDUA mRNA (iduronidase, Ref Seq: NM_000203) are provided in the table below:

Target	spacer sequence	ADAR extension sequence
CFTR	gttcatagggatccaagtttt (SEQ ID NO: 43)	tttcctccactgttgcaaag (SEQ ID NO: 44)
IDUA	ccagcgcccaccgcccccag (SEQ ID NO: 45)	acttcggcccagagctgctcc (SEQ ID NO: 46)

[0170] In one embodiment, the system disclosed herein comprises nucleic acid sequences which are minimalized to a nucleotide length which fits in a single vector. In some embodiments, the vector is an AAV vector. AAV vectors are capable of packaging transgenes which are about 4.5kbs in size. In some instances, AAV vectors are capable of packaging larger transgenes such as about 4.6 kb, 4.7 kb, 4.8 kb, 4.9 kb, 5.0 kb, 5.1 kb, 5.2 kb, 5.3 kb, 5.4 kb, 5.5 kb, 5.6 kb, 5.7 kb, 5.8 kb, 5.9 kb, 6.0 kb, 6.1 kb, 6.2 kb, 6.3 kb, 6.4 kb, 6.5 kb, 6.6 kb, 6.7 kb, 6.8 kb, 6.9 kb, 7.0 kb, 7.5 kb, 8.0 kb, 9.0 kb, 10.0 kb, 11.0 kb, 12.0 kb, 13.0 kb, 14.0 kb, 15.0 kb, or larger are used.

[0171] In another embodiment, the system disclosed herein comprises, without limitation, one or more promoter sequences for driving expression of the system components. Exemplary promoters for expressing small RNAs, without limitation, are polymerase III promoters such as U6 and H1. Other promoters for driving expression of system components are, without limitation, EF1alpha (or its short, intron-less form, EFS), CAG (CMV enhancer, chicken beta-Actin promoter and rabbit beta-Globin splice acceptor site fusion), mini CMV (cytomegalovirus), CMV, MCK (muscle creatin kinase), MCK/SV40, desmin, and/or c512 (Glutamate carboxypeptidase II).

[0172] In one embodiment, the recombinant expression system is encoded in DNA carried by a vector, e.g., adeno-associated virus (AAV), and can be delivered to appropriate tissues via one of the following methods: use of specific AAV serotypes that display specific tissue tropism (such as AAV-9 targeting neurons or muscle); injection of naked DNA encoding the RdCas9 system into tissue such as muscle or liver; use of nanoparticles composed of lipids, polymers, or other synthetic or natural materials that carry DNA or RNA encoding the therapeutic recombinant expression system; or any of the above where the system is split between two separate viruses or DNA molecules so that: one virus encodes the dCas9 protein-ADAR fusion and the other virus encodes the sgRNA; or one virus encodes the dCas9 protein and/or the sgRNA while the other virus encodes the ADAR protein and/or the sgRNA. In embodiments in which the portions of CREDIT are encoded on separate vectors, the encoded portions of dCas9 and ADAR can interact with one another so as to form a functional dCas9 – ADAR nucleoprotein complex. Exemplary split systems can be seen in

Wright et al., Rational design of a split-Cas9 enzyme complex. PNAS 112:2984-2989 (2015), the content of which is hereby incorporated by reference in its entirety).

[0173] To use exemplary recombinant expression systems as provided herein in treatment of a human subject or animal, the vector, e.g., the AAV, system can, for example, be injected by the following methods: (1) Skeletal muscle tissue (intramuscular) at multiple sites simultaneously (relevant indication: myotonic dystrophy)—injection of 10^{11}-10^{14} GC (genome copies) per injection into major muscle group such as the abdominal muscles, biceps, deltoids, erector spinae, gastrocnemius, soleus, gluteus, hamstrings, latissimus dorsi, rhomboids, obliques, *pectoralis*, quadriceps, trapezius and/or triceps; (2) Intravenous delivery of a targeted AAV serotype such as AAV-9 or AAV-6 for muscle targeting—injection of 10^{11}-10^{14} GC per injection for a total of 10^{12}-10^{17} GC delivered; 3. Subpial spinal injection of AAV-6, AAV-9 or another serotype displaying neuronal tropism —injection of 10^{11}-10^{17} GC in a single or multiple doses; 4. Intracranial injection of AAV-6, AAV-9 or another serotype displaying neuronal tropism—injection of 10^{11}-10^{17} GC in a single or multiple doses.

[0174] In other embodiments, recombinant expression systems disclosed herein may be formulated by methods known in the art. In addition, any route of administration may be envisioned such as, e.g., by any conventional route of administration including, but not limited to oral, pulmonary, intraperitoneal (ip), intravenous (iv), intramuscular (im), subcutaneous (sc), transdermal, buccal, nasal, sublingual, ocular, rectal and vaginal. In addition, administration directly to the nervous system may include, and are not limited to, intracerebral, intraventricular, intracerebroventricular, intrathecal, intracistemal, intraspinal or peri-spinal routes of administration by delivery via intracranial or intravertebral needles or catheters with or without pump devices. Any dose or frequency of administration that provides the therapeutic effect described herein is suitable for use in the present treatment. In a particular embodiment, the subject is administered a viral vector encoding the recombinant expression system according to the disclosure by the intramuscular route. In one embodiment, the vector is an AAV vector as defined above, is an AAV9 vector. In some embodiments, the human subject may receive a single injection of the vector. Additionally, standard pharmaceutical methods can be employed to control the duration of action. These are well known in the art and include control release preparations and can include appropriate

macromolecules, for example polymers, polyesters, polyamino acids, polyvinyl, pyrolidone, ethylenevinylacetate, methyl cellulose, carboxymethyl cellulose or protamine sulfate. In addition, the pharmaceutical composition may comprise nanoparticles that contain the recombinant expression system of the present disclosure.

[0175] Also provided by this invention is a composition comprising, consisting of, or consisting essentially of one or more of a recombinant expression system, vector, cell, or viral particle as described herein and a carrier. In some embodiments, the carrier is a pharmaceutically acceptable carrier.

[0176] In some embodiments, the recombinant expression systems as disclosed herein can optionally include the additional administration of a PAMmer oligonucleotide, *i.e.*, co-administration with the disclosed systems simultaneously or sequentially of a corresponding PAMmer. Selection techniques for PAMmer oligonucleotide sequences are well known in the art and can be found for example, in WO 2015/089277, incorporated herein by reference in its entirety. Although a PAMmer may in some instances increase binding affinity of dCas9 to RNA *in vivo* as well as *in vitro*, Applicants' prior work WO 2017/091630, incorporated herein by reference in its entirety, surprisingly found that a PAMmer is not required to achieve RNA recognition and editing. To simplify Applicants' delivery strategy herein and to maintain the disclosed systems herein as fully encodeable systems, the experiments below were performed in the absence of a PAMmer. A schematic of this mechanism is outlined in FIG. 1A.

[0177] Disclosed herein are methods of using recombinant expression systems as disclosed herein as a research tool, *e.g.* to characterize the effects of directed cellular RNA editing on processing and dynamics.

[0178] Additionally disclosed herein are methods of using recombinant expression systems as disclosed herein as a therapeutic for diseases, *e.g.* by using viral (AAV) or other vector-based delivery approaches to deliver the recombinant expression systems for *in vivo* or *ex vivo* RNA editing to treat a disease in need of such editing.

[0179] Non-limiting examples of targets and related diseases include, but are not limited to, premature termination codon RNA diseases such as Hurler's syndrome, Cystic fibrosis,

Duchenne muscular dystrophy, others, as well as diseases associated with deficiencies in RNA editing such as excitotoxic neuronal disorders affiliated with under-editing of the Q/R residue of AMPA subunit GluA2. Excitotoxicity may be involved in spinal cord injury, stroke, traumatic brain injury, hearing loss (through noise overexposure or ototoxicity), and in neurodegenerative diseases of the central nervous system (CNS) such as multiple sclerosis, Alzheimer's disease, amyotrophic lateral sclerosis (ALS), Parkinson's disease, alcoholism or alcohol withdrawal and especially over-rapid benzodiazepine withdrawal, and also Huntington's disease.

Examples

[0180] The following examples are non-limiting and illustrative of procedures which can be used in various instances in carrying the disclosure into effect. Additionally, all reference disclosed herein below are incorporated by reference in their entirety.

[0181] Described below are prototypes of the recombinant expression system generated by Applicant that 1) recognize and edit a reporter mRNA construct in living cells at a base specific level and 2) reverse premature termination codon (PTC) mediated silencing of expression from eGFP reporter transcripts in living cells (see FIGS. 1C and 1D).

Example 1 – Directed editing of cellular RNA via nuclear delivery of CRISPR/Cas9

Plasmid Construction

[0182] The sequence encoding dCas9-2xNLS was cloned from pCDNA3.1-dCas9-2xNLS-EGFP (Addgene plasmid #74710). For the ADAR2-XTEN-dCas9 fusion product, the dCas9 sequence fused to an XTEN peptide linker and an ADAR2 catalytic domain (PCR amplified from human ADAR2 ORF) into a pCDNA3.1 (Invitrogen) backbone using Gibson assembly. The dCas9 moiety was removed by inverse PCR using primers flanking the dCas9-NLS sequence to generate the ADAR2-XTEN fusion. PCR-mediated site-directed mutagenesis was performed to generate the ADAR2-XTEN-dCas9 E488Q and ADAR2-XTEN E488Q mutant variants, using the ADAR2-XTEN-dCas9 and ADAR2-XTEN respectively as templates. All fusion sequences were cloned into pCDNA5/FRT/TO (Invitrogen) through

PCR amplification and restriction digestion using FastDigest HindIII and NotI (Thermo Fisher).

[0183] To construct the esgRNA backbone, sequences for mammalian Ef1a promoter, mCherry ORF, and BGH poly(A) signal were Gibson assembled into pBlueScript II SK (+) (Agilent) backbone bearing a modified sgRNA scaffold (Chen et al. 2013) driven by a U6 polymerase III promoter. Individual sgRNAs bearing a 3' extension sequences were generated by PCR amplifying the modified sgRNA scaffold using tailed primers bearing the spacer and extension sequences and Gibson assembling into the pBlueScript II SK(+)-mCherry vector downstream of the U6 promoter.

Cell lines and Transfections

[0184] Flp-In T-REX 293 were cultured in Dulbecco's modified eagle medium (DMEM) supplemented with 10% fetal bovine serum (Gibco). Cells were passaged every 3-4 days using TrypLE Express (Gibco) and maintained in a tissue culture incubator at 37 °C with 5% CO_2.

[0185] Stable, doxycycline-inducible lines were generated by seeding cells on 10cm tissue culture dished and co-transfecting at 60-70% confluency with 1 ug pCDNA5/FRT/TO bearing the ADAR2 fusion constructs along with 9 ug pOG44 (Invitrogen), which encodes the Flp recombinase using polyethylenimine (PEI). Cells were subsequently passaged to 25% confluency and selected with 5 ug/ml blasticidin and 100 ug/ml hygromycin B (Gibco) after 48 hours. Cells remained under selection until individual hygromycin-resistant colonies identified, and 8-10 colonies were picked for expansion and validation.

[0186] Prior to transfection, 0.1 x 10^6 cells were seeded onto a 24-well plate 24 hours prior to the day of transfection and pre-incubated with doxycycline at a final concentration of 1 ug/ml for 24 hours. Cells were then co-transfected with 150 ug of respective sgRNA-mCherry constructs with 350 ug of W58X mutant or WT eGFP reporter construct (generous gifts from Stafforst lab) using Lipofectamine 3000 (Invitrogen). Cells were kept under doxycycline induction for 48 hours following transfection before imaging and FACS analysis. Images were captured using a Zeiss fluorescence microscope at 20x magnification.

Flow Cytometry Analysis

[0187] Cells were dissociated with TrypLE Express using standard protocol. Cells were then resuspended in 1X DPBS (Corning) supplemented with 5% FBS, passed through a 35μm nylon cell strainer, and subjected to flow cytometry analysis using an LSRFortessa or Accuri instrument (BD). Cells were appropriately gated and analyzed for GFP (FITC) fluorescence. To normalize for transfection efficiency, individual values of percent eGFP corrected for each fusion-esgRNA pair was calculated by taking the fraction of GFP-positive cells from the W58X eGFP transfection population and dividing by the fraction of GFP-positive cells when instead transfected with the WT eGFP reporter. FACS analysis was analyzed using FlowJo software and compiled results were plotted using Graphpad Prism 6.

Discussion

[0188] In these experiments, and without limitation, the recombinant expression system described above comprises A) nucleic acid sequences encoding a nuclease-dead Cas9 (dCas9) protein fused to the catalytic deaminase domain of the human ADAR2 protein, and B) an extended single guide RNA (esgRNA) sequence driven by a U6 polymerase III promoter. The systems were delivered to the nuclei of mammalian cells with the appropriate transfection reagents and the sequences bind and edit target mRNA after forming an RCas9-RNA recognition complex. This allows for selective RNA editing in which targeted adenosine residues are deaminated to inosine to be recognized as guanosine by the cellular machinery.

[0189] The catalytically active deaminase domains (DD) described in the above systems were either wildtype human ADAR2 or human ADAR2 DD bearing a mutation (E488Q) that increases enzymatic activity and affinity for RNA substrate as compared to wildtype human ADAR2. The DD was fused to a semi-flexible XTEN peptide linker at its C-terminus, which was then fused to dCas9 at its N-terminus (**FIG. 1B**). To control for RNA-recognition independent background editing, fusion constructs lacking the dCas9 moiety were also generated (AX, AX-488Q).

[0190] The esgRNA construct was modified with a region of homology capable of near-perfect RNA-RNA base pairing with over the desired site of editing. The homology region

comprises a mismatch of the targeted adenosine, forcing an A-C mispairing and the generation of a 'pseudo-dsRNA' substrate on the target transcript (**FIG. 1A**). This generates a means of programmable RNA substrate recognition as well as simultaneous base-specific deamination. Furthermore, these modified esgRNA constructs were cloned into a vector additionally comprising a marker gene, e.g., mCherry construct driven by a separate Ef1a pol II promoter, as shown in the examples. This provided for the sorting of cells transfected with the esgRNA using flow-cytometry, and furthermore enrichment of cells with targeted RNA editing.

Example 2 – Comparison of dSpCas9 and dSaCas9 CREDIT systems

[0191] dSaCas9 is significantly smaller than dSpCas9, which provides efficiency in viral packaging. A CREDIT system was prepared comprising (1) an ADAR2(E488Q)-dSaCas9 fusion with a GSGS linker (SEQ ID NO: 12) and (2) an esgRNA with a scaffold sequence specific to SaCas9 that targets an EGFP reporter (SEQ ID NO: 11). The efficiency of mRNA editing by this system was compared to a system comprising ADAR2(E488Q)-dSpCas9, as shown in FIG. 13B. ADAR2-dSaCas9 resulted in about 30% of target cells expressing successfully edited EGFP RNA, as compared to about 20% by ADAR2-dSpCas9. Overall, this data shows successful editing by both ADAR2-dSaCas9 and ADAR2-dSpCas9.

Example 3 – Treatment of Limb-girdle muscular dystrophy -type 2B

[0192] Limb-girdle muscular dystrophy -type 2B is caused by a defect in the Dysferlin gene. By developing methods to accurately correct Dysferlin mRNA in a subject, a fully functional dysferlin protein can be expressed in patients with this disorder.

[0193] The recombinant expression systems of the present disclosure allow for simple correction of the mutant dysferlin mRNA. When combined with the disclosed AAV delivery system, these systems can be used to efficiently target every major muscle with a single intravenous administration, and provide a robust therapeutic strategy to treat muscular dystrophy. Because the AAV will ultimately be used to target skeletal muscle, an AAV with skeletal muscle tropism should be used such as AAV1, AAV6, AAV7, AAV8, or AAV9.

[0194] Viral particles are prepared as described herein. Briefly, Flp-In T-REX 293 cells are transfected vectors as described in Example 1. An esgRNA is designed to target the mutant locus within the subject's dysferlin mRNA. The esgRNA can be designed to target a mutation in one or more of the following dysferlin mRNAs: NM_001130455, NM_001130976, NM_001130977, NM_001130978, NM_001130979, NM_001130980, NM_001130981, NM_001130982, NM_001130983, NM_001130984, NM_001130985, NM_001130986, NM_001130987, or NM_003494). In some embodiments, the subject's dysferlin mRNA is sequenced prior to design of the esgRNA to confirm the presence of a correctable A point mutation. A nucleic acid encoding the esgRNA is cloned into a suitable vector. Following transfection of the packaging cells, assembled viral particles are harvested and tested for Cas9 protein expression, as well as expression of esgRNA. The packaged virus is also assayed for viral titer which should range from about 10^8 GC/mL to 10^17 GC/mL, with titer optimally of about 10^13 GC/mL. Viral titer can be assayed by western blot or by viral genome copy number by qPCR and compared to copy number standard samples.

[0195] Modified viral particles can be administered *ex vivo* or *in vitro* to muscle stem or progenitor cells from subjects with Limb-girdle muscular dystrophy -type 2B. Upon integration of the viral vectors, the modified cells are transplanted back into subject via intramuscular injection. Effectiveness of cell therapy with the cells treated with modified AAV is measured by improved muscle morphology, decreases in sarcolemmal localization of the multimeric dystrophin-glycoprotein complex and neuronal nitric-oxide synthase, as well as detection of dysferlin expression.

[0196] Alternatively, the viral particles can be administered *in vivo* to muscle tissue through, for example, localized or systemic delivery such as intramuscular injection, intraperitoneal injection, or intravenous injection. Effectiveness of viral gene therapy is measured by improved muscle morphology as well as detection of dysferlin expression.

[0197] Efficiency of CRISPR –mediated RNA editing is assayed by designing PCR primers that detect a reverse transcribed copy of the repaired dysferlin mRNA fragment. Expression of repaired gene product can also be detected by PCR, histological staining, or western blot of treated muscle tissue.

Example 4 – Editing of CFTR mRNA

[0198] Cystic fibrosis is a genetic disorder that affects the lungs, pancreas, liver, kidneys, and intestine. Long-term symptoms include difficulty breathing and coughing up mucus as a result of frequent lung infections. Other signs and symptoms may include sinus infections, poor growth, fatty stool, clubbing of the fingers and toes, and infertility. Cystic fibrosis is caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene. By developing methods to accurately correct CFTR mRNA in a subject, a fully functional CFTR protein can be expressed in these patients.

[0199] The recombinant expression systems of the present disclosure allow for simple correction of CFTR mRNA. When combined with the a viral delivery system such as AAV or lentivirus, these systems can be used to efficiently target affected tissues and provide a robust therapeutic strategy to treat Cystic Fibrosis. AAV with lung tropism include but are not limited to AAV4, AAV5, AAV6, and AAV9.

[0200] An esgRNA is designed to target the mutant locus within the subject's CTFR mRNA. In some embodiments, the subject's CFTR mRNA is sequenced prior to design of the esgRNA to confirm the presence of a correctable A point mutation. A nucleic acid encoding the esgRNA is cloned into a suitable vector. A non-limiting example of a suitable CFTR targeting spacer sequence is SEQ ID NO: 43. A non-limiting example of a suitable CFTR extension sequence is SEQ ID NO: 44. A non-limiting example of a lentiviral plasmid comprising an esgRNA targeted to CFTR is LCV2_purpo_CFTR_51_1217_gibson (SEQ ID NO: 35).

[0201] Following transfection of the packaging cells, assembled viral particles are harvested and tested for Cas9 protein expression, as well as expression of esgRNA. The packaged virus is also assayed for viral titer which should range from about 10^8 GC/mL to 10^17 GC/mL, with titer optimally of about 10^13 GC/mL. Viral titer can be assayed by western blot or by viral genome copy number by qPCR and compared to copy number standard samples.

[0202] Viral particles can be administered *in vivo* to the subject through, for example, localized or systemic delivery such as intraperitoneal injection, organ-targeted injection, or

intravenous injection. Effectiveness of viral gene therapy is measured by improved lung function, a reduction or amelioration of one or more symptoms of Cystic Fibrosis, and/or detection of corrected CFTR protein expression.

[0203] Efficiency of CRISPR –mediated RNA editing is assayed by designing PCR primers that detect a reverse transcribed copy of the repaired CFTR mRNA fragment. Expression of repaired gene product can also be detected by PCR, histological staining, or western blot of treated lung tissue.

Example 5 – Editing of IDUA mRNA

[0204] Hurler syndrome is a genetic disorder that results in the buildup of glycosaminoglycans due to a deficiency of alpha-L iduronidase (IDUA), an enzyme responsible for the degradation of mucopolysaccharides in lysosomes. Without this enzyme, a buildup of dermatan sulfate and heparan sulphate occurs in the body. Symptoms include but are not limited to hepatosplenomegaly, dwarfism, unique facial features, progressive mental retardation, and early death due to organ damage.

[0205] The recombinant expression systems of the present disclosure allow for simple correction of IDUA mRNA. When combined with the a viral delivery system such as AAV or lentivirus, these systems can be used to provide a robust therapeutic strategy to treat Hurler syndrome.

[0206] An esgRNA is designed to target the mutant locus within the subject's IDUA mRNA. In some embodiments, the subject's IDUA mRNA is sequenced prior to design of the esgRNA to confirm the presence of a correctable A point mutation. A nucleic acid encoding the esgRNA is cloned into a suitable vector. A non-limiting example of a suitable IDUA targeting spacer sequence is SEQ ID NO: 45. A non-limiting example of a suitable IDUA extension sequence is SEQ ID NO: 46. A non-limiting example of a lentiviral plasmid comprising an esgRNA targeted to IDUA is AXCM_LCV2_puro_IDUA_No-spacer_gibson (SEQ ID NO: 39).

[0207] Following transfection of the packaging cells, assembled viral particles are harvested and tested for Cas9 protein expression, as well as expression of esgRNA. The

packaged virus is also assayed for viral titer which should range from about 10^8 GC/mL to 10^17 GC/mL, with titer optimally of about 10^13 GC/mL. Viral titer can be assayed by western blot or by viral genome copy number by qPCR and compared to copy number standard samples.

[0208] Viral particles can be administered *in vivo* to the subject through, for example, systemic delivery such as intravenous injection. Effectiveness of viral gene therapy is measured by decrease in the amount of heparin sulphate in the subject, a reduction or amelioration of one or more symptoms of Hurler syndrome, and/or detection of corrected IDUA protein expression.

[0209] Efficiency of CRISPR –mediated RNA editing is assayed by designing PCR primers that detect a reverse transcribed copy of the repaired IDUA mRNA fragment. Expression of repaired gene product can also be detected by PCR, histological staining, or western blot of treated tissues.

Equivalents

[0210] It should be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification, improvement and variation of the inventions embodied therein herein disclosed may be resorted to by those skilled in the art, and that such modifications, improvements and variations are considered to be with the scope of this invention. The materials, methods, and examples provided here are representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the invention.

[0211] The invention has been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the generic disclosure also form part of the invention. This includes the generic description of the invention with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein.

[0212] In addition, where features or aspects of the invention are described in terms of Markush groups, those skilled in the art will recognize that the invention is also thereby

described in terms of any individual member or subgroup of members of the Markush group.

[0213] All publications, patent applications, patents, and other references mentioned herein are expressly incorporated by reference in their entirety, to the same extent as if each were incorporated by reference individually. In case of conflict, the present specification, including definitions, will control.

References

1. Fukuda, M., et al., Construction of a guide-RNA for site-directed RNA mutagenesis utilising intracellular A-to-I RNA editing. Sci Rep, 2017. 7: p. 41478.

2. Halo et al "NanoFlares for the detection, isolation, and culture of live tumor cells from human blood" PNAS doi: 10.1073/pnas.1418637111.

3. Hanswillemenke et al., Site-Directed RNA Editing in Vivo Can Be Triggered by the Light-Driven Assembly of an Artificial Riboprotein. J Am Chem Soc, 2015. 137(50): p. 15875-81.

4. Hua et al "Peripheral SMN restoration is essential for long-term rescue of a severe spinal muscular atrophy mouse model." Nature. 2011 Oct 5;478(7367):123-6. doi: 10.1038/nature10485.

5. McMahon et al., TRIBE: Hijacking an RNA-Editing Enzyme to Identify Cell-Specific Targets of RNA-Binding Proteins. Cell, 2016. 165(3): p. 742-53.

6. Montiel-Gonzalez et al "An efficient system for selectively altering genetic information within mRNAs." Nucleic Acids Res. 2016 44: e157. doi: 10.1093/nar/gkw738.

7. Montiel-Gonzalez et al "Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA editing." PNAS. 2013 110: 18285-90.

8. Schneider et al "Optimal guideRNAs for re-directing deaminase activity of hADAR1 and hADAR2 in *trans*." Nucleic Acids Res. 2014 42: e87. doi: 10.1093/nar/gku272.

9. Wang et al "Engineering splicing factors with designed specificities" Nat Methods. 2009 Nov; 6(11): 825–830. 10.1038/nmeth.1379

10. WO 2015089277

11. WO 2016183402

Sequences

[0214] Provided below are exemplary sequences of the constructs described herein.

pcDNA3.1(1)_ADAR2_XTEN_dCas9 (SEQ ID NO: 27)

LOCUS Exported 10826 bp ds-DNA circular
DEFINITION synthetic circular DNA
SOURCE synthetic DNA construct
ORGANISM recombinant plasmid
REFERENCE 1 (bases 1 to 10826)
FEATURES Location/Qualifiers
source 1..10826
/organism="recombinant plasmid"
/mol_type="other DNA"
enhancer 235..614
/label=CMV enhancer
/note="human cytomegalovirus immediate early enhancer"
promoter 615..818
/label=CMV promoter
/note="human cytomegalovirus (CMV) immediate early promoter"
promoter 863..881
/label=T7 promoter
/note="promoter for bacteriophage T7 RNA polymerase"
misc_feature 927..954
/label=Homology 1_pCDNA3.1
primer_bind 955..976
/label=ADAR2CD-Cas9_HindIII_F
misc_feature 955..960
/label=Kozak
primer_bind 960..983
/label=Adar_out_forward_lv2
CDS 961..2100
/codon_start=1
/label=ADARB1_Catalytic Domain

/translation="MLADAVSRLVLGKFGDLTDNFSSPHARRKVLAGVVMTTGTDVKDA
KVISVSTGTKCINGEYMSDRGLALNDCHAEIISRRSLLRFLYTQLELYLNNKDDQKRSI
FQKSERGGFRLKENVQFHLYISTSPCGDARIFSPHEPILEEPADRHPNRKARGQLRTKI
ESGEGTIPVRSNASIQTWDGVLQGERLLTMSCSDKIARWNVVGIQGSLLSIFVEPIYFS

SIILGSLYHGDHLSRAMYQRISNIEDLPPLYTLNKPLLSGISNAEARQPGKAPNFSVNW
TVGDSAIEVINATTGKDELGRASRLCKHALYCRWMRVHGKVPSHLLRSKITKPNVYHES
KLAAKEYQAAKARLFTAFIKAGLGAWVEKPTEQDQFSLTP"
primer_bind 1324..1346
/label=E488Q_ADAR2_Mut_seq
primer_bind complement(1426..1447)
/label=E488Q_Mut_Classic_R
primer_bind 1448..1472
/label=E488Q_Mut_Classic_F
CDS 2101..2148
/codon_start=1
/label=XTEN
/translation="SGSETPGTSESATPES"
primer_bind complement(2129..2148)
/label=ADAR2_CD_Inverse_R
CDS 2149..6252
/codon_start=1
/product="catalytically dead mutant of the Cas9
endonuclease from the Streptococcus pyogenes Type II
CRISPR/Cas system"
/label=dCas9
/note="RNA-guided DNA-binding protein that lacks
endonuclease activity due to the D10A mutation in the RuvC
catalytic domain and the H840A mutation in the HNH
catalytic domain"

/translation="MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKK
NLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEES
FLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIK
FRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL
ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQ
IGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLL
RKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARG
NSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER
LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFM
QLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGR
HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY
LYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVP

SEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKT
EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSK
ESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL
DATLIHQSITGLYETRIDLSQLGGD"

primer_bind complement(6233..6252)
/label=Cas9_out_rev_lv2

primer_bind 6253..6274
/label=ADAR2_CD_Inverse_F

CDS 6256..6282
/codon_start=1
/product="HA (human influenza hemagglutinin) epitope tag"
/label=HA
/translation="YPYDVPDYA"

CDS 6301..6321
/codon_start=1
/product="nuclear localization signal of SV40 large T antigen"
/label=SV40 NLS
/translation="PKKKRKV"

CDS 6328..6348
/codon_start=1
/product="nuclear localization signal of SV40 large T antigen"
/label=SV40 NLS
/translation="PKKKRKV"

primer_bind complement(6332..6357)
/label=ADAR2CD-Cas9_NotI_R

misc_feature 6358..6392
/label=Homology 2_pCDNA3.1

polyA_signal 6426..6650
/label=bGH poly(A) signal
/note="bovine growth hormone polyadenylation signal"

rep_origin 6696..7124
/direction=RIGHT
/label=f1 ori
/note="f1 bacteriophage origin of replication; arrow

indicates direction of (+) strand synthesis"
promoter 7138..7467
/label=SV40 promoter
/note="SV40 enhancer and early promoter"
rep_origin 7318..7453
/label=SV40 ori
/note="SV40 origin of replication"
CDS 7534..8328
/codon_start=1
/gene="aph(3')-II (or nptII)"
/product="aminoglycoside phosphotransferase from Tn5"
/label=NeoR/KanR
/note="confers resistance to neomycin, kanamycin, and G418
(Geneticin(R))"

/translation="MIEQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQGRP
VLFVKTDLSGALNELQDEAARLSWLATTGVPCAAVLDVVTEAGRDWLLLGEVPGQDLLS
SHLAPAEKVSIMADAMRRLHTLDPATCPFDHQAKHRIERARTRMEAGLVDQDDLDEEHQ
GLAPAELFARLKARMPDGEDLVVTHGDACLPNIMVENGRFSGFIDCGRLGVADRYQDIA
LATRDIAEELGGEWADRFLVLYGIAAPDSQRIAFYRLLDEFF"
polyA_signal 8502..8623
/label=SV40 poly(A) signal
/note="SV40 polyadenylation signal"
primer_bind complement(8672..8688)
/label=M13 rev
/note="common sequencing primer, one of multiple similar
variants"
protein_bind 8696..8712
/label=lac operator
/bound_moiety="lac repressor encoded by lacI"
/note="The lac repressor binds to the lac operator to
inhibit transcription in E. coli. This inhibition can be
relieved by adding lactose or
isopropyl-beta-D-thiogalactopyranoside (IPTG)."
promoter complement(8720..8750)
/label=lac promoter
/note="promoter for the E. coli lac operon"
protein_bind 8765..8786
/label=CAP binding site
/bound_moiety="E. coli catabolite activator protein"
/note="CAP binding activates transcription in the presence

of cAMP."
rep_origin complement(9074..9659)
/direction=LEFT
/label=ori
/note="high-copy-number ColE1/pMB1/pBR322/pUC origin of replication"
CDS complement(9830..10690)
/codon_start=1
/gene="bla"
/product="beta-lactamase"
/label=AmpR
/note="confers resistance to ampicillin, carbenicillin, and related antibiotics"

/translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGYI
ELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVEYS
PVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRLDRW
EPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSA
LPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGAS
LIKHW"
promoter complement(10691..10795)
/gene="bla"
/label=AmpR promoter
ORIGIN
1 gacggatcgg gagatctccc gatcccctat ggtgcactct cagtacaatc tgctctgatg
61 ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgct gagtagtgcg
121 cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc
181 ttagggttag gcgttttgcg ctgcttcgcg atgtacgggc cagatatacg cgttgacatt
241 gattattgac tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata
301 tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc
361 cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc
421 attgacgtca atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt
481 atcatatgcc aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt
541 atgcccagta catgacctta tgggactttc ctacttggca gtacatctac gtattagtca
601 tcgctattac catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg
661 actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc
721 aaaatcaacg ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg
781 gtaggcgtgt acggtgggag gtctatataa gcagagctct ctggctaact agagaaccca
841 ctgcttactg gcttatcgaa attaatacga ctcactatag ggagacccaa gctggctagc
901 gtttaaacgg gccctctaga ctcgagcggc cgccactgtg ctggatatct gcagagaacc
961 atgttagctg acgctgtctc acgcctggtc ctgggtaagt ttggtgacct gaccgacaac

1021 ttctcctccc ctcacgctcg cagaaaagtg ctggctggag tcgtcatgac aacaggcaca
1081 gatgttaaag atgccaaggt gataagtgtt tctacaggaa caaaatgtat taatggtgaa
1141 tacatgagtg atcgtggcct tgcattaaat gactgccatg cagaaataat atctcggaga
1201 tccttgctca gatttcttta tacacaactt gagctttact taaataacaa agatgatcaa
1261 aaaagatcca tctttcagaa atcagagcga ggggggttta ggctgaagga gaatgtccag
1321 tttcatctgt acatcagcac ctctccctgt ggagatgcca gaatcttctc accacatgag
1381 ccaatcctgg aagaaccagc agatagacac ccaaatcgta aagcaagagg acagctacgg
1441 accaaaatag agtctggtga ggggacgatt ccagtgcgct ccaatgcgag catccaaacg
1501 tgggacgggg tgctgcaagg ggagcggctg ctcaccatgt cctgcagtga caagattgca
1561 cgctggaacg tggtgggcat ccagggatcc ctgctcagca ttttcgtgga gcccatttac
1621 ttctcgagca tcatcctggg cagcctttac cacggggacc acctttccag ggccatgtac
1681 cagcggatct ccaacataga ggacctgcca cctctctaca ccctcaacaa gcctttgctc
1741 agtggcatca gcaatgcaga agcacggcag ccagggaagg cccccaactt cagtgtcaac
1801 tggacggtag gcgactccgc tattgaggtc atcaacgcca cgactgggaa ggatgagctg
1861 ggccgcgcgt cccgcctgtg taagcacgcg ttgtactgtc gctggatgcg tgtgcacggc
1921 aaggttccct cccacttact acgctccaag attaccaagc ccaacgtgta ccatgagtcc
1981 aagctggcgg caaaggagta ccaggccgcc aaggcgcgtc tgttcacagc cttcatcaag
2041 gcggggctgg gggcctgggt ggagaagccc accgagcagg accagttctc actcacgccc
2101 agtggaagtg agacaccggg aacctcagag agcgccacgc cagaaagcat ggacaagaag
2161 tacagcatcg gcctggccat cggcaccaac tctgtgggct gggccgtgat caccgacgag
2221 tacaaggtgc ccagcaagaa attcaaggtg ctgggcaaca ccgaccggca cagcatcaag
2281 aagaacctga tcggcgccct gctgttcgac agcggagaaa cagccgaggc cacccggctg
2341 aagagaaccg ccagaagaag atacaccaga cggaagaacc ggatctgcta tctgcaagag
2401 atcttcagca acgagatggc caaggtggac gacagcttct tccacagact ggaagagtcc
2461 ttcctggtgg aagaggataa gaagcacgag cggcacccca tcttcggcaa catcgtggac
2521 gaggtggcct accacgagaa gtaccccacc atctaccacc tgagaaagaa actggtggac
2581 agcaccgaca aggccgacct gcggctgatc tatctggccc tggcccacat gatcaagttc
2641 cggggccact tcctgatcga gggcgacctg aaccccgaca acagcgacgt ggacaagctg
2701 ttcatccagc tggtgcagac ctacaaccag ctgttcgagg aaaaccccat caacgccagc
2761 ggcgtggacg ccaaggccat cctgtctgcc agactgagca agagcagacg gctggaaaat
2821 ctgatcgccc agctgcccgg cgagaagaag aatggcctgt tcggcaacct gattgccctg
2881 agcctgggcc tgaccccca cttcaagagc aacttcgacc tggccgagga tgccaaactg
2941 cagctgagca aggacaccta cgacgacgac ctggacaacc tgctggccca gatcggcgac
3001 cagtacgccg acctgtttct ggccgccaag aacctgtccg acgccatcct gctgagcgac
3061 atcctgagag tgaacaccga gatcaccaag gcccccctga gcgcctctat gatcaagaga
3121 tacgacgagc accaccagga cctgaccctg ctgaaagctc tcgtgcggca gcagctgcct
3181 gagaagtaca aagagatttt cttcgaccag agcaagaacg gctacgccgg ctacatcgat
3241 ggcggagcca gccaggaaga gttctacaag ttcatcaagc ccatcctgga aaagatggac
3301 ggcaccgagg aactgctcgt gaagctgaac agagaggacc tgctgcggaa gcagcggacc
3361 ttcgacaacg gcagcatccc ccaccagatc cacctgggag agctgcacgc cattctgcgg
3421 cggcaggaag atttttaccc attcctgaag gacaaccggg aaaagatcga gaagatcctg

3481 accttccgca tcccctacta cgtgggccct ctggccaggg gaaacagcag attcgcctgg
3541 atgaccagaa agagcgagga aaccatcacc ccctggaact tcgaggaagt ggtggacaag
3601 ggcgccagcg cccagagctt catcgagcgg atgaccaact tcgataagaa cctgcccaac
3661 gagaaggtgc tgcccaagca cagcctgctg tacgagtact tcaccgtgta caacgagctg
3721 accaaagtga aatacgtgac cgagggaatg agaaagcccg ccttcctgag cggcgagcag
3781 aaaaaagcca tcgtggacct gctgttcaag accaaccgga aagtgaccgt gaagcagctg
3841 aaagaggact acttcaagaa aatcgagtgc ttcgactccg tggaaatctc cggcgtggaa
3901 gatcggttca acgcctccct gggcacatac cacgatctgc tgaaaattat caaggacaag
3961 gacttcctgg acaatgagga aaacgaggac attctggaag atatcgtgct gaccctgaca
4021 ctgtttgagg acagagagat gatcgaggaa cggctgaaaa cctatgccca cctgttcgac
4081 gacaaagtga tgaagcagct gaagcggcgg agatacaccg gctggggcag gctgagccgg
4141 aagctgatca acggcatccg ggacaagcag tccggcaaga caatcctgga tttcctgaag
4201 tccgacggct tcgccaacag aaacttcatg cagctgatcc acgacgacag cctgaccttt
4261 aaagaggaca tccagaaagc ccaggtgtcc ggccagggcg atagcctgca cgagcacatt
4321 gccaatctgg ccggcagccc cgccattaag aagggcatcc tgcagacagt gaaggtggtg
4381 gacgagctcg tgaaagtgat gggccggcac aagcccgaga acatcgtgat cgaaatggcc
4441 agagagaacc agaccaccca gaagggacag aagaacagcc gcgagagaat gaagcggatc
4501 gaagagggca tcaaagagct gggcagccag atcctgaaag aacaccccgt ggaaaacacc
4561 cagctgcaga acgagaagct gtacctgtac tacctgcaga atgggcggga tatgtacgtg
4621 gaccaggaac tggacatcaa ccggctgtcc gactacgatg tggacgctat cgtgcctcag
4681 agctttctga aggacgactc catcgataac aaagtgctga ctcggagcga caagaaccgg
4741 ggcaagagcg acaacgtgcc ctccgaagag gtcgtgaaga agatgaagaa ctactggcgc
4801 cagctgctga atgccaagct gattacccag aggaagttcg acaatctgac caaggccgag
4861 agaggcggcc tgagcgaact ggataaggcc ggcttcatca agagacagct ggtggaaacc
4921 cggcagatca caaagcacgt ggcacagatc ctggactccc ggatgaacac taagtacgac
4981 gagaacgaca aactgatccg ggaagtgaaa gtgatcaccc tgaagtccaa gctggtgtcc
5041 gatttccgga aggatttcca gttttacaaa gtgcgcgaga tcaacaacta ccaccacgcc
5101 cacgacgcct acctgaacgc cgtcgtggga accgccctga tcaaaaagta ccctaagctg
5161 gaaagcgagt tcgtgtacgg cgactacaag gtgtacgacg tgcggaagat gatcgccaag
5221 agcgagcagg aaatcggcaa ggctaccgcc aagtacttct tctacagcaa catcatgaac
5281 tttttcaaga ccgagattac cctggccaac ggcgagatcc ggaagcggcc tctgatcgag
5341 acaaacggcg aaacaggcga gatcgtgtgg gataagggcc gggactttgc caccgtgcgg
5401 aaagtgctgt ctatgcccca agtgaatatc gtgaaaaaga ccgaggtgca gacaggcggc
5461 ttcagcaaag agtctatcct gcccaagagg aacagcgaca agctgatcgc cagaaagaag
5521 gactgggacc ctaagaagta cggcggcttc gacagcccca ccgtggccta ttctgtgctg
5581 gtggtggcca aagtggaaaa gggcaagtcc aagaaactga agagtgtgaa agagctgctg
5641 gggatcacca tcatggaaag aagcagcttc gagaagaatc ccatcgactt tctggaagcc
5701 aagggctaca aagaagtgaa aaaggacctg atcatcaagc tgcctaagta ctccctgttc
5761 gagctggaaa acggccggaa gagaatgctg gcctctgccg gcgaactgca gaagggaaac
5821 gaactggccc tgccctccaa atatgtgaac ttcctgtacc tggccagcca ctatgagaag
5881 ctgaagggct cccccgagga taatgagcag aaacagctgt ttgtggaaca gcacaaacac

5941 tacctggacg agatcatcga gcagatcagc gagttctcca agagagtgat cctggccgac
6001 gctaatctgg acaaggtgct gagcgcctac aacaagcaca gagacaagcc tatcagagag
6061 caggccgaga atatcatcca cctgtttacc ctgaccaatc tgggagcccc tgccgccttc
6121 aagtactttg acaccaccat cgaccggaag aggtacacca gcaccaaaga ggtgctggac
6181 gccaccctga tccaccagag catcaccggc ctgtacgaga cacggatcga cctgtctcag
6241 ctgggaggcg acgcctatcc ctatgacgtg cccgattatg ccagcctggg cagcggctcc
6301 cccaagaaaa aacgcaaggt ggaagatcct aagaaaaagc ggaaagtgga cgtgtaacca
6361 ccacactgga ctagtggatc cgagctcggt accaagctta agtttaaacc gctgatcagc
6421 ctcgactgtg ccttctagtt gccagccatc tgttgtttgc ccctcccccg tgccttcctt
6481 gaccctggaa ggtgccactc ccactgtcct ttcctaataa aatgaggaaa ttgcatcgca
6541 ttgtctgagt aggtgtcatt ctattctggg gggtggggtg gggcaggaca gcaaggggga
6601 ggattgggaa gacaatagca ggcatgctgg ggatgcggtg ggctctatgg cttctgaggc
6661 ggaaagaacc agctggggct ctaggggta tccccacgcg ccctgtagcg gcgcattaag
6721 cgcggcgggt gtggtggtta cgcgcagcgt gaccgctaca cttgccagcg ccctagcgcc
6781 cgctcctttc gctttcttcc cttcctttct cgccacgttc gccggctttc cccgtcaagc
6841 tctaaatcgg gggctccctt tagggttccg atttagtgct ttacggcacc tcgaccccaa
6901 aaaacttgat tagggtgatg gttcacgtag tgggccatcg ccctgataga cggtttttcg
6961 ccctttgacg ttggagtcca cgttctttaa tagtggactc ttgttccaaa ctggaacaac
7021 actcaaccct atctcggtct attcttttga tttataaggg attttgccga tttcggccta
7081 ttggttaaaa aatgagctga tttaacaaaa atttaacgcg aattaattct gtggaatgtg
7141 tgtcagttag ggtgtggaaa gtccccaggc tccccagcag gcagaagtat gcaaagcatg
7201 catctcaatt agtcagcaac caggtgtgga aagtccccag gctccccagc aggcagaagt
7261 atgcaaagca tgcatctcaa ttagtcagca accatagtcc cgcccctaac tccgcccatc
7321 ccgcccctaa ctccgcccag ttccgcccat tctccgcccc atggctgact aatttttttt
7381 atttatgcag aggccgaggc cgcctctgcc tctgagctat tccagaagta gtgaggaggc
7441 ttttttggag gcctaggctt ttgcaaaaag ctcccgggag cttgtatatc cattttcgga
7501 tctgatcaag agacaggatg aggatcgttt cgcatgattg aacaagatgg attgcacgca
7561 ggttctccgg ccgcttgggt ggagaggcta ttcggctatg actgggcaca acagacaatc
7621 ggctgctctg atgccgccgt gttccggctg tcagcgcagg ggcgcccggt tctttttgtc
7681 aagaccgacc tgtccggtgc cctgaatgaa ctgcaggacg aggcagcgcg gctatcgtgg
7741 ctggccacga cgggcgttcc ttgcgcagct gtgctcgacg ttgtcactga agcgggaagg
7801 gactggctgc tattgggcga agtgccgggg caggatctcc tgtcatctca ccttgctcct
7861 gccgagaaag tatccatcat ggctgatgca atgcggcggc tgcatacgct tgatccggct
7921 acctgcccat tcgaccacca agcgaaacat cgcatcgagc gagcacgtac tcggatggaa
7981 gccggtcttg tcgatcagga tgatctggac gaagagcatc aggggctcgc gccagccgaa
8041 ctgttcgcca ggctcaaggc gcgcatgccc gacggcgagg atctcgtcgt gacccatggc
8101 gatgcctgct tgccgaatat catggtggaa aatggccgct tttctggatt catcgactgt
8161 ggccggctgg gtgtggcgga ccgctatcag gacatagcgt tggctacccg tgatattgct
8221 gaagagcttg gcggcgaatg ggctgaccgc ttcctcgtgc tttacggtat cgccgctccc
8281 gattcgcagc gcatcgcctt ctatcgcctt cttgacgagt tcttctgagc gggactctgg
8341 ggttcgaaat gaccgaccaa gcgacgccca acctgccatc acgagatttc gattccaccg

8401 ccgccttcta tgaaaggttg ggcttcggaa tcgttttccg ggacgccggc tggatgatcc
8461 tccagcgcgg ggatctcatg ctggagttct tcgcccaccc caacttgttt attgcagctt
8521 ataatggtta caaataaagc aatagcatca caaatttcac aaataaagca ttttttcac
8581 tgcattctag ttgtggtttg tccaaactca tcaatgtatc ttatcatgtc tgtataccgt
8641 cgacctctag ctagagcttg gcgtaatcat ggtcatagct gtttcctgtg tgaaattgtt
8701 atccgctcac aattccacac aacatacgag ccggaagcat aaagtgtaaa gcctggggtg
8761 cctaatgagt gagctaactc acattaattg cgttgcgctc actgcccgct ttccagtcgg
8821 gaaacctgtc gtgccagctg cattaatgaa tcggccaacg cgcggggaga ggcggtttgc
8881 gtattgggcg ctcttccgct tcctcgctca ctgactcgct gcgctcggtc gttcggctgc
8941 ggcgagcggt atcagctcac tcaaaggcgg taatacggtt atccacagaa tcaggggata
9001 acgcaggaaa gaacatgtga gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg
9061 cgttgctggc gtttttccat aggctccgcc cccctgacga gcatcacaaa aatcgacgct
9121 caagtcagag gtggcgaaac ccgacaggac tataaagata ccaggcgttt ccccctggaa
9181 gctccctcgt gcgctctcct gttccgaccc tgccgcttac cggatacctg tccgcctttc
9241 tcccttcggg aagcgtggcg ctttctcata gctcacgctg taggtatctc agttcggtgt
9301 aggtcgttcg ctccaagctg ggctgtgtgc acgaaccccc cgttcagccc gaccgctgcg
9361 ccttatccgg taactatcgt cttgagtcca acccggtaag acacgactta tcgccactgg
9421 cagcagccac tggtaacagg attagcagag cgaggtatgt aggcggtgct acagagttct
9481 tgaagtggtg gcctaactac ggctacacta gaagaacagt atttggtatc tgcgctctgc
9541 tgaagccagt taccttcgga aaaagagttg gtagctcttg atccggcaaa caaaccaccg
9601 ctggtagcgg tttttttgtt tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag
9661 aagatccttt gatcttttct acggggtctg acgctcagtg gaacgaaaac tcacgttaag
9721 ggattttggt catgagatta tcaaaaagga tcttcaccta gatccttta aattaaaaat
9781 gaagttttaa atcaatctaa agtatatatg agtaaacttg gtctgacagt taccaatgct
9841 taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcatccata gttgcctgac
9901 tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa
9961 tgataccgcg agacccacgc tcaccggctc cagatttatc agcaataaac cagccagccg
10021 gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt
10081 gttgccggga agctagagta agtagttcgc cagttaatag tttgcgcaac gttgttgcca
10141 ttgctacagg catcgtggtg tcacgctcgt cgtttggtat ggcttcattc agctccggtt
10201 cccaacgatc aaggcgagtt acatgatccc ccatgttgtg caaaaaagcg gttagctcct
10261 tcggtcctcc gatcgttgtc agaagtaagt tggccgcagt gttatcactc atggttatgg
10321 cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct gtgactggtg
10381 agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg
10441 cgtcaatacg ggataatacc gcgccacata gcagaacttt aaaagtgctc atcattggaa
10501 aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt
10561 aacccactcg tgcacccaac tgatcttcag catcttttac tttcaccagc gtttctgggt
10621 gagcaaaaac aggaaggcaa aatgccgcaa aaaagggaat aagggcgaca cggaaatgtt
10681 gaatactcat actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca
10741 tgagcggata catatttgaa tgtatttaga aaaataaaca aatagggttc cgcgcacat
10801 ttccccgaaa agtgccacct gacgtc

pcDNA3.1(1)_ADAR2_XTEN_control (SEQ ID NO: 28).

LOCUS Exported 6722 bp ds-DNA circular
DEFINITION synthetic circular DNA
FEATURES Location/Qualifiers
source 1..6722
/organism="synthetic DNA construct"
/mol_type="other DNA"
enhancer 235..614
/label=CMV enhancer
/note="human cytomegalovirus immediate early enhancer"
promoter 615..818
/label=CMV promoter
/note="human cytomegalovirus (CMV) immediate early promoter"
promoter 863..881
/label=T7 promoter
/note="promoter for bacteriophage T7 RNA polymerase"
misc_feature 927..954
/label=Homology 1_pCDNA3.1
primer_bind 955..976
/label=ADAR2CD-Cas9_HindIII_F
primer_bind 960..983
/label=Adar_out_forward_lv2
CDS 961..2100
/codon_start=1
/label=ADARB1(E488Q)_Catalytic Domain

/translation="MLADAVSRLVLGKFGDLTDNFSSPHARRKVLAGVVMTTGTDVKDA
KVISVSTGTKCINGEYMSDRGLALNDCHAEIISRRSLLRFLYTQLELYLNNKDDQKRSI
FQKSERGGFRLKENVQFHLYISTSPCGDARIFSPHEPILEEPADRHPNRKARGQLRTKI
ESGEGTIPVRSNASIQTWDGVLQGERLLTMSCSDKIARWNVVGIQGSLLSIFVEPIYFS
SIILGSLYHGDHLSRAMYQRISNIEDLPPLYTLNKPLLSGISNAEARQPGKAPNFSVNW
TVGDSAIEVINATTGKDELGRASRLCKHALYCRWMRVHGKVPSHLLRSKITKPNVYHES
KLAAKEYQAAKARLFTAFIKAGLGAWVEKPTEQDQFSLTP"
primer_bind 1324..1346
/label=E488Q_ADAR2_Mut
primer_bind complement(1426..1447)
/label=E488Q_Mut_Classic_R
primer_bind 1448..1472
/label=E488Q_Mut_Classic_F

CDS 2101..2148
/codon_start=1
/label=XTEN
/translation="SGSETPGTSESATPES"
primer_bind complement(2129..2148)
/label=ADAR2_CD_Inverse_R
primer_bind 2149..2170
/label=ADAR2_CD_Inverse_F
CDS 2152..2178
/codon_start=1
/product="HA (human influenza hemagglutinin) epitope tag"
/label=HA
/translation="YPYDVPDYA"
CDS 2197..2217
/codon_start=1
/product="nuclear localization signal of SV40 large T
antigen"
/label=SV40 NLS
/translation="PKKKRKV"
CDS 2224..2244
/codon_start=1
/product="nuclear localization signal of SV40 large T
antigen"
/label=SV40 NLS
/translation="PKKKRKV"
primer_bind complement(2228..2253)
/label=ADAR2CD-Cas9_NotI_R
misc_feature 2254..2288
/label=Homology 2_pCDNA3.1
polyA_signal 2322..2546
/label=bGH poly(A) signal
/note="bovine growth hormone polyadenylation signal"
rep_origin 2592..3020
/direction=RIGHT
/label=f1 ori
/note="f1 bacteriophage origin of replication; arrow
indicates direction of (+) strand synthesis"
promoter 3034..3363
/label=SV40 promoter
/note="SV40 enhancer and early promoter"
rep_origin 3214..3349

/label=SV40 ori
/note="SV40 origin of replication"
CDS 3430..4224
/codon_start=1
/gene="aph(3')-II (or nptII)"
/product="aminoglycoside phosphotransferase from Tn5"
/label=NeoR/KanR
/note="confers resistance to neomycin, kanamycin, and G418
(Geneticin(R))"

/translation="MIEQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQGRP
VLFVKTDLSGALNELQDEAARLSWLATTGVPCAAVLDVVTEAGRDWLLLGEVPGQDLLS
SHLAPAEKVSIMADAMRRLHTLDPATCPFDHQAKHRIERARTRMEAGLVDQDDLDEEHQ
GLAPAELFARLKARMPDGEDLVVTHGDACLPNIMVENGRFSGFIDCGRLGVADRYQDIA
LATRDIAEELGGEWADRFLVLYGIAAPDSQRIAFYRLLDEFF"
polyA_signal 4398..4519
/label=SV40 poly(A) signal
/note="SV40 polyadenylation signal"
primer_bind complement(4568..4584)
/label=M13 rev
/note="common sequencing primer, one of multiple similar
variants"
protein_bind 4592..4608
/label=lac operator
/bound_moiety="lac repressor encoded by lacI"
/note="The lac repressor binds to the lac operator to
inhibit transcription in E. coli. This inhibition can be
relieved by adding lactose or
isopropyl-beta-D-thiogalactopyranoside (IPTG)."
promoter complement(4616..4646)
/label=lac promoter
/note="promoter for the E. coli lac operon"
protein_bind 4661..4682
/label=CAP binding site
/bound_moiety="E. coli catabolite activator protein"
/note="CAP binding activates transcription in the presence
of cAMP."
rep_origin complement(4970..5555)
/direction=LEFT
/label=ori
/note="high-copy-number ColE1/pMB1/pBR322/pUC origin of

replication"
CDS complement(5726..6586)
/codon_start=1
/gene="bla"
/product="beta-lactamase"
/label=AmpR
/note="confers resistance to ampicillin, carbenicillin, and related antibiotics"

/translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGYI
ELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVEYS
PVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRLDRW
EPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSA
LPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGAS
LIKHW" promoter complement(6587..6691)
/gene="bla"
/label=AmpR promoter
ORIGIN
1 gacggatcgg gagatctccc gatcccctat ggtgcactct cagtacaatc tgctctgatg
61 ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgct gagtagtgcg
121 cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc
181 ttagggttag gcgttttgcg ctgcttcgcg atgtacgggc cagatatacg cgttgacatt
241 gattattgac tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata
301 tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc
361 cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc
421 attgacgtca atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt
481 atcatatgcc aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt
541 atgcccagta catgacctta tgggactttc ctacttggca gtacatctac gtattagtca
601 tcgctattac catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg
661 actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc
721 aaaatcaacg ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg
781 gtaggcgtgt acggtgggag gtctatataa gcagagctct ctggctaact agagaaccca
841 ctgcttactg gcttatcgaa attaatacga ctcactatag ggagacccaa gctggctagc
901 gtttaaacgg gccctctaga ctcgagcggc cgccactgtg ctggatatct gcagagaacc
961 atgttagctg acgctgtctc acgcctggtc ctgggtaagt ttggtgacct gaccgacaac
1021 ttctcctccc ctcacgctcg cagaaaagtg ctggctggag tcgtcatgac aacaggcaca
1081 gatgttaaag atgccaaggt gataagtgtt tctacaggaa caaaatgtat taatggtgaa
1141 tacatgagtg atcgtggcct tgcattaaat gactgccatg cagaaataat atctcggaga
1201 tccttgctca gatttcttta tacacaactt gagctttact taaataacaa agatgatcaa
1261 aaaagatcca tctttcagaa atcagagcga ggggggttta ggctgaagga gaatgtccag
1321 tttcatctgt acatcagcac ctctccctgt ggagatgcca gaatcttctc accacatgag

1381 ccaatcctgg aagaaccagc agatagacac ccaaatcgta aagcaagagg acagctacgg
1441 accaaaatag agtctggtga ggggacgatt ccagtgcgct ccaatgcgag catccaaacg
1501 tgggacgggg tgctgcaagg ggagcggctg ctcaccatgt cctgcagtga caagattgca
1561 cgctggaacg tggtgggcat ccagggatcc ctgctcagca ttttcgtgga gcccatttac
1621 ttctcgagca tcatcctggg cagcctttac cacggggacc acctttccag ggccatgtac
1681 cagcggatct ccaacataga ggacctgcca cctctctaca ccctcaacaa gcctttgctc
1741 agtggcatca gcaatgcaga agcacggcag ccagggaagg cccccaactt cagtgtcaac
1801 tggacggtag gcgactccgc tattgaggtc atcaacgcca cgactgggaa ggatgagctg
1861 ggccgcgcgt cccgcctgtg taagcacgcg ttgtactgtc gctggatgcg tgtgcacggc
1921 aaggttccct cccacttact acgctccaag attaccaagc ccaacgtgta ccatgagtcc
1981 aagctggcgg caaaggagta ccaggccgcc aaggcgcgtc tgttcacagc cttcatcaag
2041 gcggggctgg gggcctgggt ggagaagccc accgagcagg accagttctc actcacgccc
2101 agtggaagtg agacaccggg aacctcagag agcgccacgc cagaaagcgc ctatccctat
2161 gacgtgcccg attatgccag cctgggcagc ggctccccca agaaaaaacg caaggtggaa
2221 gatcctaaga aaaagcggaa agtggacgtg taaccaccac actggactag tggatccgag
2281 ctcggtacca agcttaagtt taaaccgctg atcagcctcg actgtgcctt ctagttgcca
2341 gccatctgtt gtttgcccct cccccgtgcc ttccttgacc ctggaaggtg ccactcccac
2401 tgtcctttcc taataaaatg aggaaattgc atcgcattgt ctgagtaggt gtcattctat
2461 tctgggggggt ggggtggggc aggacagcaa gggggaggat tgggaagaca atagcaggca
2521 tgctggggat gcggtgggct ctatggcttc tgaggcggaa agaaccagct ggggctctag
2581 ggggtatccc cacgcgccct gtagcggcgc attaagcgcg gcgggtgtgg tggttacgcg
2641 cagcgtgacc gctacacttg ccagcgccct agcgcccgct cctttcgctt tcttcccttc
2701 ctttctcgcc acgttcgccg gctttccccg tcaagctcta aatcgggggc tccctttagg
2761 gttccgattt agtgctttac ggcacctcga ccccaaaaaa cttgattagg gtgatggttc
2821 acgtagtggg ccatcgccct gatagacggt ttttcgccct ttgacgttgg agtccacgtt
2881 ctttaatagt ggactcttgt tccaaactgg aacaacactc aaccctatct cggtctattc
2941 ttttgattta taagggattt tgccgatttc ggcctattgg ttaaaaaatg agctgattta
3001 acaaaaattt aacgcgaatt aattctgtgg aatgtgtgtc agttagggtg tggaaagtcc
3061 ccaggctccc cagcaggcag aagtatgcaa agcatgcatc tcaattagtc agcaaccagg
3121 tgtggaaagt ccccaggctc cccagcaggc agaagtatgc aaagcatgca tctcaattag
3181 tcagcaacca tagtcccgcc cctaactccg cccatcccgc ccctaactcc gcccagttcc
3241 gcccattctc cgccccatgg ctgactaatt ttttttattt atgcagaggc cgaggccgcc
3301 tctgcctctg agctattcca gaagtagtga ggaggctttt ttggaggcct aggcttttgc
3361 aaaaagctcc cgggagcttg tatatccatt ttcggatctg atcaagagac aggatgagga
3421 tcgtttcgca tgattgaaca agatggattg cacgcaggtt ctccggccgc ttgggtggag
3481 aggctattcg gctatgactg ggcacaacag acaatcggct gctctgatgc cgccgtgttc
3541 cggctgtcag cgcaggggcg cccggttctt tttgtcaaga ccgacctgtc cggtgccctg
3601 aatgaactgc aggacgaggc agcgcggcta tcgtggctgg ccacgacggg cgttccttgc
3661 gcagctgtgc tcgacgttgt cactgaagcg ggaagggact ggctgctatt gggcgaagtg
3721 ccggggcagg atctcctgtc atctcacctt gctcctgccg agaaagtatc catcatggct
3781 gatgcaatgc ggcggctgca tacgcttgat ccggctacct gcccattcga ccaccaagcg

3841 aaacatcgca tcgagcgagc acgtactcgg atggaagccg gtcttgtcga tcaggatgat
3901 ctggacgaag agcatcaggg gctcgcgcca gccgaactgt tcgccaggct caaggcgcgc
3961 atgcccgacg gcgaggatct cgtcgtgacc catggcgatg cctgcttgcc gaatatcatg
4021 gtggaaaatg gccgcttttc tggattcatc gactgtggcc ggctgggtgt ggcggaccgc
4081 tatcaggaca tagcgttggc tacccgtgat attgctgaag agcttggcgg cgaatgggct
4141 gaccgcttcc tcgtgcttta cggtatcgcc gctcccgatt cgcagcgcat cgccttctat
4201 cgccttcttg acgagttctt ctgagcggga ctctggggtt cgaaatgacc gaccaagcga
4261 cgcccaacct gccatcacga gatttcgatt ccaccgccgc cttctatgaa aggttgggct
4321 tcggaatcgt tttccgggac gccggctgga tgatcctcca gcgcggggat ctcatgctgg
4381 agttcttcgc ccaccccaac ttgtttattg cagcttataa tggttacaaa taaagcaata
4441 gcatcacaaa tttcacaaat aaagcatttt tttcactgca ttctagttgt ggtttgtcca
4501 aactcatcaa tgtatcttat catgtctgta taccgtcgac ctctagctag agcttggcgt
4561 aatcatggtc atagctgttt cctgtgtgaa attgttatcc gctcacaatt ccacacaaca
4621 tacgagccgg aagcataaag tgtaaagcct ggggtgccta atgagtgagc taactcacat
4681 taattgcgtt gcgctcactg cccgctttcc agtcgggaaa cctgtcgtgc cagctgcatt
4741 aatgaatcgg ccaacgcgcg gggagaggcg gtttgcgtat tgggcgctct tccgcttcct
4801 cgctcactga ctcgctgcgc tcggtcgttc ggctgcggcg agcggtatca gctcactcaa
4861 aggcggtaat acggttatcc acagaatcag gggataacgc aggaaagaac atgtgagcaa
4921 aaggccagca aaaggccagg aaccgtaaaa aggccgcgtt gctggcgttt ttccataggc
4981 tccgcccccc tgacgagcat cacaaaaatc gacgctcaag tcagaggtgg cgaaacccga
5041 caggactata aagataccag gcgtttcccc ctggaagctc cctcgtgcgc tctcctgttc
5101 cgaccctgcc gcttaccgga tacctgtccg cctttctccc ttcgggaagc gtggcgcttt
5161 ctcatagctc acgctgtagg tatctcagtt cggtgtaggt cgttcgctcc aagctgggct
5221 gtgtgcacga acccccccgtt cagcccgacc gctgcgcctt atccggtaac tatcgtcttg
5281 agtccaaccc ggtaagacac gacttatcgc cactggcagc agccactggt aacaggatta
5341 gcagagcgag gtatgtaggc ggtgctacag agttcttgaa gtggtggcct aactacggct
5401 acactagaag aacagtattt ggtatctgcg ctctgctgaa gccagttacc ttcggaaaaa
5461 gagttggtag ctcttgatcc ggcaaacaaa ccaccgctgg tagcggtttt tttgtttgca
5521 agcagcagat tacgcgcaga aaaaaaggat ctcaagaaga tcctttgatc ttttctacgg
5581 ggtctgacgc tcagtggaac gaaaactcac gttaagggat tttggtcatg agattatcaa
5641 aaaggatctt cacctagatc cttttaaatt aaaaatgaag ttttaaatca atctaaagta
5701 tatatgagta aacttggtct gacagttacc aatgcttaat cagtgaggca cctatctcag
5761 cgatctgtct atttcgttca tccatagttg cctgactccc cgtcgtgtag ataactacga
5821 tacgggaggg cttaccatct ggccccagtg ctgcaatgat accgcgagac ccacgctcac
5881 cggctccaga tttatcagca ataaaccagc cagccggaag ggccgagcgc agaagtggtc
5941 ctgcaacttt atccgcctcc atccagtcta ttaattgttg ccgggaagct agagtaagta
6001 gttcgccagt taatagtttg cgcaacgttg ttgccattgc tacaggcatc gtggtgtcac
6061 gctcgtcgtt tggtatggct tcattcagct ccggttccca acgatcaagg cgagttacat
6121 gatcccccat gttgtgcaaa aaagcggtta gctccttcgg tcctccgatc gttgtcagaa
6181 gtaagttggc cgcagtgtta tcactcatgg ttatggcagc actgcataat tctcttactg
6241 tcatgccatc cgtaagatgc ttttctgtga ctggtgagta ctcaaccaag tcattctgag

6301 aatagtgtat gcggcgaccg agttgctctt gcccggcgtc aatacgggat aataccgcgc
6361 cacatagcag aactttaaaa gtgctcatca ttggaaaacg ttcttcgggg cgaaaactct
6421 caaggatctt accgctgttg agatccagtt cgatgtaacc cactcgtgca cccaactgat
6481 cttcagcatc ttttactttc accagcgttt ctgggtgagc aaaaacagga aggcaaaatg
6541 ccgcaaaaaa gggaataagg gcgacacgga aatgttgaat actcatactc ttccttttc
6601 aatattattg aagcatttat cagggttatt gtctcatgag cggatacata tttgaatgta
6661 tttagaaaaa taaacaaata ggggttccgc gcacatttcc ccgaaaagtg ccacctgacg
6721 tc

pcDNA3.1_ADAR2(E488Q)_XTEN_dCas9 (SEQ ID NO: 29).

LOCUS Exported 10826 bp ds-DNA circular
DEFINITION synthetic circular DNA
SOURCE synthetic DNA construct
ORGANISM synthetic DNA construct
REFERENCE 1 (bases 1 to 10826)
FEATURES Location/Qualifiers
source 1..10826
/organism="synthetic DNA construct"
/mol_type="other DNA"
enhancer 235..614
/label=CMV enhancer
/note="human cytomegalovirus immediate early enhancer"
promoter 615..818
/label=CMV promoter
/note="human cytomegalovirus (CMV) immediate early promoter"
promoter 863..881
/label=T7 promoter
/note="promoter for bacteriophage T7 RNA polymerase"
primer_bind 927..985
/label=H1-ADAR-XTEN_F
misc_feature 927..954
/label=Homology 1_pCDNA3.1
CDS 961..2100
/codon_start=1
/label=ADARB1(E488Q)_Catalytic Domain

/translation="MLADAVSRLVLGKFGDLTDNFSSPHARRKVLAGVVMTTGTDVKDA
KVISVSTGTKCINGEYMSDRGLALNDCHAEIISRRSLLRFLYTQLELYLNNKDDQKRSI

FQKSERGGFRLKENVQFHLYISTSPCGDARIFSPHEPILEEPADRHPNRKARGQLRTKI
ESGQGTIPVRSNASIQTWDGVLQGERLLTMSCSDKIARWNVVGIQGSLLSIFVEPIYFS
SIILGSLYHGDHLSRAMYQRISNIEDLPPLYTLNKPLLSGISNAEARQPGKAPNFSVNW
TVGDSAIEVINATTGKDELGRASRLCKHALYCRWMRVHGKVPSHLLRSKITKPNVYHES
KLAAKEYQAAKARLFTAFIKAGLGAWVEKPTEQDQFSLTP"

primer_bind 961..982
/label=Primer 4
primer_bind 1111..1138
/label=Primer 1
primer_bind 1440..1478
/label=E488Q_Mutagenesis_F
primer_bind complement(1440..1478)
/label=E488Q_Mutagenesis_R
primer_bind complement(2080..2100)
/label=ADAR2DD_GS_R
primer_bind complement(2080..2100)
/label=Primer 5
CDS 2101..2148
/codon_start=1
/label=XTEN
/translation="SGSETPGTSESATPES"
primer_bind complement(2129..2148)
/label=ADAR2_XTEN_R
primer_bind complement(2129..2148)
/label=ADAR2_CD_Inverse_R
primer_bind 2148..2171
/label=Primer 2
CDS 2149..6252
/codon_start=1
/product="catalytically dead mutant of the Cas9
endonuclease from the Streptococcus pyogenes Type II
CRISPR/Cas system"
/label=dCas9
/note="RNA-guided DNA-binding protein that lacks
endonuclease activity due to the D10A mutation in the RuvC
catalytic domain and the H840A mutation in the HNH
catalytic domain"

/translation="MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKK
NLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEES
FLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIK

FRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL
ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQ
IGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLL
RKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARG
NSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER
LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFM
QLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGR
HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY
LYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVP
SEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKT
EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSK
ESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL
DATLIHQSITGLYETRIDLSQLGGD"

primer_bind complement(4458..4479)
/label=Primer 3
primer_bind 4879..4899
/label=Primer 6
primer_bind 6252..6273
/label=SaCas9_HA_F
primer_bind 6253..6274
/label=ADAR2_CD_Inverse_F
CDS 6256..6282
/codon_start=1
/product="HA (human influenza hemagglutinin) epitope tag"
/label=HA
/translation="YPYDVPDYA"
primer_bind complement(6274..6296)
/label=AXC_NLSout_NESin_R
primer_bind complement(6274..6294)
/label=NLS_out_R
CDS 6301..6321
/codon_start=1
/product="nuclear localization signal of SV40 large T

antigen"
/label=SV40 NLS
/translation="PKKKRKV"
CDS 6328..6348
/codon_start=1
/product="nuclear localization signal of SV40 large T
antigen"
/label=SV40 NLS
/translation="PKKKRKV"
primer_bind complement(6333..6392)
/label=XTEN-Cas9-H2_R
primer_bind complement(6333..6377)
/label=Primer 7
primer_bind 6347..6371
/label=NLS_out_NES_full_F
primer_bind 6349..6371
/label=AXC_NLSout_NESin_F
misc_feature 6358..6392
/label=Homology 2_pCDNA3.1
polyA_signal 6426..6650
/label=bGH poly(A) signal
/note="bovine growth hormone polyadenylation signal"
rep_origin 6696..7124
/direction=RIGHT
/label=f1 ori
/note="f1 bacteriophage origin of replication; arrow
indicates direction of (+) strand synthesis"
promoter 7138..7467
/label=SV40 promoter
/note="SV40 enhancer and early promoter"
rep_origin 7318..7453
/label=SV40 ori
/note="SV40 origin of replication"
CDS 7534..8328
/codon_start=1
/gene="aph(3')-II (or nptII)"
/product="aminoglycoside phosphotransferase from Tn5"
/label=NeoR/KanR
/note="confers resistance to neomycin, kanamycin, and G418
(Geneticin(R))"

/translation="MIEQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQGRP
VLFVKTDLSGALNELQDEAARLSWLATTGVPCAAVLDVVTEAGRDWLLLGEVPGQDLLS
SHLAPAEKVSIMADAMRRLHTLDPATCPFDHQAKHRIERARTRMEAGLVDQDDLDEEHQ
GLAPAELFARLKARMPDGEDLVVTHGDACLPNIMVENGRFSGFIDCGRLGVADRYQDIA
LATRDIAEELGGEWADRFLVLYGIAAPDSQRIAFYRLLDEFF"

polyA_signal 8502..8623
/label=SV40 poly(A) signal
/note="SV40 polyadenylation signal"

primer_bind complement(8672..8688)
/label=M13 rev
/note="common sequencing primer, one of multiple similar variants"

protein_bind 8696..8712
/label=lac operator
/bound_moiety="lac repressor encoded by lacI"
/note="The lac repressor binds to the lac operator to inhibit transcription in E. coli. This inhibition can be relieved by adding lactose or isopropyl-beta-D-thiogalactopyranoside (IPTG)."

promoter complement(8720..8750)
/label=lac promoter
/note="promoter for the E. coli lac operon"

protein_bind 8765..8786
/label=CAP binding site
/bound_moiety="E. coli catabolite activator protein"
/note="CAP binding activates transcription in the presence of cAMP."

rep_origin complement(9074..9659)
/direction=LEFT
/label=ori
/note="high-copy-number ColE1/pMB1/pBR322/pUC origin of replication"

CDS complement(9830..10690)
/codon_start=1
/gene="bla"
/product="beta-lactamase"
/label=AmpR
/note="confers resistance to ampicillin, carbenicillin, and related antibiotics"

/translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGYI
ELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVEYS
PVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRLDRW
EPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSA
LPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGAS
LIKHW"

promoter complement(10691..10795)

/gene="bla"

/label=AmpR promoter

ORIGIN

1 gacggatcgg gagatctccc gatcccctat ggtgcactct cagtacaatc tgctctgatg
61 ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgct gagtagtgcg
121 cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc
181 ttagggttag gcgttttgcg ctgcttcgcg atgtacgggc cagatatacg cgttgacatt
241 gattattgac tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata
301 tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc
361 cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc
421 attgacgtca atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt
481 atcatatgcc aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt
541 atgcccagta catgacctta tgggactttc ctacttggca gtacatctac gtattagtca
601 tcgctattac catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg
661 actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc
721 aaaatcaacg ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg
781 gtaggcgtgt acggtgggag gtctatataa gcagagctct ctggctaact agagaaccca
841 ctgcttactg gcttatcgaa attaatacga ctcactatag ggagacccaa gctggctagc
901 gtttaaacgg gccctctaga ctcgagcggc cgccactgtg ctggatatct gcagagaacc
961 atgttagctg acgctgtctc acgcctggtc ctgggtaagt ttggtgacct gaccgacaac
1021 ttctcctccc ctcacgctcg cagaaaagtg ctggctggag tcgtcatgac aacaggcaca
1081 gatgttaaag atgccaaggt gataagtgtt tctacaggaa caaaatgtat taatggtgaa
1141 tacatgagtg atcgtggcct tgcattaaat gactgccatg cagaaataat atctcggaga
1201 tccttgctca gatttcttta tacacaactt gagctttact taaataacaa agatgatcaa
1261 aaaagatcca tctttcagaa atcagagcga ggggggttta ggctgaagga gaatgtccag
1321 tttcatctgt acatcagcac ctctccctgt ggagatgcca gaatcttctc accacatgag
1381 ccaatcctgg aagaaccagc agatagacac ccaaatcgta aagcaagagg acagctacgg
1441 accaaaatag agtctggtca ggggacgatt ccagtgcgct ccaatgcgag catccaaacg
1501 tgggacgggg tgctgcaagg ggagcggctg ctcaccatgt cctgcagtga caagattgca
1561 cgctggaacg tggtgggcat ccagggatcc ctgctcagca ttttcgtgga gcccatttac
1621 ttctcgagca tcatcctggg cagcctttac cacggggacc acctttccag ggccatgtac
1681 cagcggatct ccaacataga ggacctgcca cctctctaca ccctcaacaa gcctttgctc
1741 agtggcatca gcaatgcaga agcacggcag ccagggaagg cccccaactt cagtgtcaac

1801 tggacggtag gcgactccgc tattgaggtc atcaacgcca cgactgggaa ggatgagctg
1861 ggccgcgcgt cccgcctgtg taagcacgcg ttgtactgtc gctggatgcg tgtgcacggc
1921 aaggttccct cccacttact acgctccaag attaccaagc ccaacgtgta ccatgagtcc
1981 aagctggcgg caaaggagta ccaggccgcc aaggcgcgtc tgttcacagc cttcatcaag
2041 gcggggctgg gggcctgggt ggagaagccc accgagcagg accagttctc actcacgccc
2101 agtggaagtg agacaccggg aacctcagag agcgccacgc cagaaagcat ggacaagaag
2161 tacagcatcg gcctggccat cggcaccaac tctgtgggct gggccgtgat caccgacgag
2221 tacaaggtgc ccagcaagaa attcaaggtg ctgggcaaca ccgaccggca cagcatcaag
2281 aagaacctga tcggcgccct gctgttcgac agcggagaaa cagccgaggc cacccggctg
2341 aagagaaccg ccagaagaag atacaccaga cggaagaacc ggatctgcta tctgcaagag
2401 atcttcagca acgagatggc caaggtggac gacagcttct tccacagact ggaagagtcc
2461 ttcctggtgg aagaggataa gaagcacgag cggcacccca tcttcggcaa catcgtggac
2521 gaggtggcct accacgagaa gtaccccacc atctaccacc tgagaaagaa actggtggac
2581 agcaccgaca aggccgacct gcggctgatc tatctggccc tggcccacat gatcaagttc
2641 cggggccact tcctgatcga gggcgacctg aaccccgaca acagcgacgt ggacaagctg
2701 ttcatccagc tggtgcagac ctacaaccag ctgttcgagg aaaaccccat caacgccagc
2761 ggcgtggacg ccaaggccat cctgtctgcc agactgagca agagcagacg gctggaaaat
2821 ctgatcgccc agctgccccgg cgagaagaag aatggcctgt tcggcaacct gattgccctg
2881 agcctgggcc tgaccccaa cttcaagagc aacttcgacc tggccgagga tgccaaactg
2941 cagctgagca aggacaccta cgacgacgac ctggacaacc tgctggccca gatcggcgac
3001 cagtacgccg acctgtttct ggccgccaag aacctgtccg acgccatcct gctgagcgac
3061 atcctgagag tgaacaccga gatcaccaag gccccctga gcgcctctat gatcaagaga
3121 tacgacgagc accaccagga cctgaccctg ctgaaagctc tcgtgcggca gcagctgcct
3181 gagaagtaca aagagatttt cttcgaccag agcaagaacg gctacgccgg ctacatcgat
3241 ggcggagcca gccaggaaga gttctacaag ttcatcaagc ccatcctgga aaagatggac
3301 ggcaccgagg aactgctcgt gaagctgaac agagaggacc tgctgcggaa gcagcggacc
3361 ttcgacaacg gcagcatccc ccaccagatc cacctgggag agctgcacgc cattctgcgg
3421 cggcaggaag atttttaccc attcctgaag gacaaccggg aaaagatcga gaagatcctg
3481 accttccgca tcccctacta cgtgggccct ctggccaggg gaaacagcag attcgcctgg
3541 atgaccagaa agagcgagga aaccatcacc ccctggaact tcgaggaagt ggtggacaag
3601 ggcgccagcg cccagagctt catcgagcgg atgaccaact tcgataagaa cctgcccaac
3661 gagaaggtgc tgcccaagca cagcctgctg tacgagtact tcaccgtgta caacgagctg
3721 accaaagtga aatacgtgac cgagggaatg agaaagcccg ccttcctgag cggcgagcag
3781 aaaaaagcca tcgtggacct gctgttcaag accaaccgga aagtgaccgt gaagcagctg
3841 aaagaggact acttcaagaa aatcgagtgc ttcgactccg tggaaatctc cggcgtggaa
3901 gatcggttca acgcctccct gggcacatac cacgatctgc tgaaaattat caaggacaag
3961 gacttcctgg acaatgagga aaacgaggac attctggaag atatcgtgct gaccctgaca
4021 ctgtttgagg acagagagat gatcgaggaa cggctgaaaa cctatgccca cctgttcgac
4081 gacaaagtga tgaagcagct gaagcggcgg agatacaccg gctggggcag gctgagccgg
4141 aagctgatca acggcatccg ggacaagcag tccggcaaga caatcctgga tttcctgaag
4201 tccgacggct tcgccaacag aaacttcatg cagctgatcc acgacgacag cctgaccttt

4261 aaagaggaca tccagaaagc ccaggtgtcc ggccagggcg atagcctgca cgagcacatt
4321 gccaatctgg ccggcagccc cgccattaag aagggcatcc tgcagacagt gaaggtggtg
4381 gacgagctcg tgaaagtgat gggccggcac aagcccgaga acatcgtgat cgaaatggcc
4441 agagagaacc agaccaccca gaagggacag aagaacagcc gcgagagaat gaagcggatc
4501 gaagagggca tcaaagagct gggcagccag atcctgaaag aacaccccgt ggaaaacacc
4561 cagctgcaga acgagaagct gtacctgtac tacctgcaga atgggcggga tatgtacgtg
4621 gaccaggaac tggacatcaa ccggctgtcc gactacgatg tggacgctat cgtgcctcag
4681 agctttctga aggacgactc catcgataac aaagtgctga ctcggagcga caagaaccgg
4741 ggcaagagcg acaacgtgcc ctccgaagag gtcgtgaaga agatgaagaa ctactggcgc
4801 cagctgctga atgccaagct gattacccag aggaagttcg acaatctgac caaggccgag
4861 agaggcggcc tgagcgaact ggataaggcc ggcttcatca agagacagct ggtggaaacc
4921 cggcagatca caaagcacgt ggcacagatc ctggactccc ggatgaacac taagtacgac
4981 gagaacgaca aactgatccg ggaagtgaaa gtgatcaccc tgaagtccaa gctggtgtcc
5041 gatttccgga aggatttcca gttttacaaa gtgcgcgaga tcaacaacta ccaccacgcc
5101 cacgacgcct acctgaacgc cgtcgtggga accgccctga tcaaaaagta ccctaagctg
5161 gaaagcgagt tcgtgtacgg cgactacaag gtgtacgacg tgcggaagat gatcgccaag
5221 agcgagcagg aaatcggcaa ggctaccgcc aagtacttct tctacagcaa catcatgaac
5281 tttttcaaga ccgagattac cctggccaac ggcgagatcc ggaagcggcc tctgatcgag
5341 acaaacggcg aaacaggcga gatcgtgtgg gataagggcc gggactttgc caccgtgcgg
5401 aaagtgctgt ctatgcccca agtgaatatc gtgaaaaaga ccgaggtgca gacaggcggc
5461 ttcagcaaag agtctatcct gcccaagagg aacagcgaca agctgatcgc cagaaagaag
5521 gactgggacc ctaagaagta cggcggcttc gacagcccca ccgtggccta ttctgtgctg
5581 gtggtggcca aagtggaaaa gggcaagtcc aagaaactga agagtgtgaa agagctgctg
5641 gggatcacca tcatggaaag aagcagcttc gagaagaatc ccatcgactt tctggaagcc
5701 aagggctaca aagaagtgaa aaaggacctg atcatcaagc tgcctaagta ctccctgttc
5761 gagctggaaa acggccggaa gagaatgctg gcctctgccg gcgaactgca gaagggaaac
5821 gaactggccc tgccctccaa atatgtgaac ttcctgtacc tggccagcca ctatgagaag
5881 ctgaagggct cccccgagga taatgagcag aaacagctgt ttgtggaaca gcacaaacac
5941 tacctggacg agatcatcga gcagatcagc gagttctcca agagagtgat cctggccgac
6001 gctaatctgg acaaggtgct gagcgcctac aacaagcaca gagacaagcc tatcagagag
6061 caggccgaga atatcatcca cctgtttacc ctgaccaatc tgggagcccc tgccgccttc
6121 aagtactttg acaccaccat cgaccggaag aggtacacca gcaccaaaga ggtgctggac
6181 gccaccctga tccaccagag catcaccggc ctgtacgaga cacggatcga cctgtctcag
6241 ctgggaggcg acgcctatcc ctatgacgtg cccgattatg ccagcctggg cagcggctcc
6301 cccaagaaaa aacgcaaggt ggaagatcct aagaaaaagc ggaaagtgga cgtgtaacca
6361 ccacactgga ctagtggatc cgagctcggt accaagctta agtttaaacc gctgatcagc
6421 ctcgactgtg ccttctagtt gccagccatc tgttgtttgc ccctcccccg tgccttcctt
6481 gaccctggaa ggtgccactc ccactgtcct ttcctaataa aatgaggaaa ttgcatcgca
6541 ttgtctgagt aggtgtcatt ctattctggg gggtggggtg gggcaggaca gcaaggggga
6601 ggattgggaa gacaatagca ggcatgctgg ggatgcggtg ggctctatgg cttctgaggc
6661 ggaaagaacc agctggggct ctaggggtga tccccacgcg ccctgtagcg gcgcattaag

6721 cgcggcgggt gtggtggtta cgcgcagcgt gaccgctaca cttgccagcg ccctagcgcc
6781 cgctcctttc gctttcttcc cttcctttct cgccacgttc gccggctttc cccgtcaagc
6841 tctaaatcgg gggctccctt tagggttccg atttagtgct ttacggcacc tcgaccccaa
6901 aaaacttgat tagggtgatg gttcacgtag tgggccatcg ccctgataga cggtttttcg
6961 ccctttgacg ttggagtcca cgttctttaa tagtggactc ttgttccaaa ctggaacaac
7021 actcaaccct atctcggtct attcttttga tttataaggg attttgccga tttcggccta
7081 ttggttaaaa aatgagctga tttaacaaaa atttaacgcg aattaattct gtggaatgtg
7141 tgtcagttag ggtgtggaaa gtccccaggc tccccagcag gcagaagtat gcaaagcatg
7201 catctcaatt agtcagcaac caggtgtgga aagtccccag gctccccagc aggcagaagt
7261 atgcaaagca tgcatctcaa ttagtcagca accatagtcc cgcccctaac tccgcccatc
7321 ccgcccctaa ctccgcccag ttccgcccat tctccgcccc atggctgact aatttttttt
7381 atttatgcag aggccgaggc cgcctctgcc tctgagctat tccagaagta gtgaggaggc
7441 ttttttggag gcctaggctt ttgcaaaaag ctcccgggag cttgtatatc cattttcgga
7501 tctgatcaag agacaggatg aggatcgttt cgcatgattg aacaagatgg attgcacgca
7561 ggttctccgg ccgcttgggt ggagaggcta ttcggctatg actgggcaca acagacaatc
7621 ggctgctctg atgccgccgt gttccggctg tcagcgcagg ggcgcccggt tctttttgtc
7681 aagaccgacc tgtccggtgc cctgaatgaa ctgcaggacg aggcagcgcg gctatcgtgg
7741 ctggccacga cgggcgttcc ttgcgcagct gtgctcgacg ttgtcactga agcgggaagg
7801 gactggctgc tattgggcga agtgccgggg caggatctcc tgtcatctca ccttgctcct
7861 gccgagaaag tatccatcat ggctgatgca atgcggcggc tgcatacgct tgatccggct
7921 acctgcccat tcgaccacca agcgaaacat cgcatcgagc gagcacgtac tcggatggaa
7981 gccggtcttg tcgatcagga tgatctggac gaagagcatc aggggctcgc gccagccgaa
8041 ctgttcgcca ggctcaaggc gcgcatgccc gacggcgagg atctcgtcgt gacccatggc
8101 gatgcctgct tgccgaatat catggtggaa aatggccgct tttctggatt catcgactgt
8161 ggccggctgg gtgtggcgga ccgctatcag gacatagcgt tggctacccg tgatattgct
8221 gaagagcttg gcggcgaatg ggctgaccgc ttcctcgtgc tttacggtat cgccgctccc
8281 gattcgcagc gcatcgcctt ctatcgcctt cttgacgagt tcttctgagc gggactctgg
8341 ggttcgaaat gaccgaccaa gcgacgccca acctgccatc acgagatttc gattccaccg
8401 ccgccttcta tgaaaggttg ggcttcggaa tcgttttccg ggacgccggc tggatgatcc
8461 tccagcgcgg ggatctcatg ctggagttct tcgcccaccc caacttgttt attgcagctt
8521 ataatggtta caaataaagc aatagcatca caaatttcac aaataaagca ttttttcac
8581 tgcattctag ttgtggtttg tccaaactca tcaatgtatc ttatcatgtc tgtataccgt
8641 cgacctctag ctagagcttg gcgtaatcat ggtcatagct gtttcctgtg tgaaattgtt
8701 atccgctcac aattccacac aacatacgag ccggaagcat aaagtgtaaa gcctggggtg
8761 cctaatgagt gagctaactc acattaattg cgttgcgctc actgcccgct ttccagtcgg
8821 gaaacctgtc gtgccagctg cattaatgaa tcggccaacg cgcggggaga ggcggtttgc
8881 gtattgggcg ctcttccgct tcctcgctca ctgactcgct gcgctcggtc gttcggctgc
8941 ggcgagcggt atcagctcac tcaaaggcgg taatacggtt atccacagaa tcaggggata
9001 acgcaggaaa gaacatgtga gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg
9061 cgttgctggc gtttttccat aggctccgcc cccctgacga gcatcacaaa aatcgacgct
9121 caagtcagag gtggcgaaac ccgacaggac tataaagata ccaggcgttt ccccctggaa

9181 gctccctcgt gcgctctcct gttccgaccc tgccgcttac cggatacctg tccgcctttc
9241 tcccttcggg aagcgtggcg ctttctcata gctcacgctg taggtatctc agttcggtgt
9301 aggtcgttcg ctccaagctg ggctgtgtgc acgaaccccc cgttcagccc gaccgctgcg
9361 ccttatccgg taactatcgt cttgagtcca acccggtaag acacgactta tcgccactgg
9421 cagcagccac tggtaacagg attagcagag cgaggtatgt aggcggtgct acagagttct
9481 tgaagtggtg gcctaactac ggctacacta gaagaacagt atttggtatc tgcgctctgc
9541 tgaagccagt taccttcgga aaaagagttg gtagctcttg atccggcaaa caaaccaccg
9601 ctggtagcgg tttttttgtt tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag
9661 aagatccttt gatcttttct acggggtctg acgctcagtg gaacgaaaac tcacgttaag
9721 ggattttggt catgagatta tcaaaaagga tcttcaccta gatcctttta aattaaaaat
9781 gaagttttaa atcaatctaa agtatatatg agtaaacttg gtctgacagt taccaatgct
9841 taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcatccata gttgcctgac
9901 tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa
9961 tgataccgcg agacccacgc tcaccggctc cagatttatc agcaataaac cagccagccg
10021 gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt
10081 gttgccggga agctagagta agtagttcgc cagttaatag tttgcgcaac gttgttgcca
10141 ttgctacagg catcgtggtg tcacgctcgt cgtttggtat ggcttcattc agctccggtt
10201 cccaacgatc aaggcgagtt acatgatccc ccatgttgtg caaaaaagcg gttagctcct
10261 tcggtcctcc gatcgttgtc agaagtaagt tggccgcagt gttatcactc atggttatgg
10321 cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct gtgactggtg
10381 agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg
10441 cgtcaatacg ggataatacc gcgccacata gcagaacttt aaaagtgctc atcattggaa
10501 aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt
10561 aacccactcg tgcacccaac tgatcttcag catcttttac tttcaccagc gtttctgggt
10621 gagcaaaaac aggaaggcaa aatgccgcaa aaaagggaat aagggcgaca cggaaatgtt
10681 gaatactcat actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca
10741 tgagcggata catatttgaa tgtatttaga aaaataaaca aatagggggtt ccgcgcacat
10801 ttccccgaaa agtgccacct gacgtc

pcDNA3.1_ADAR2(E488Q)_XTEN_control (SEQ ID NO: 30).

LOCUS Exported 6722 bp ds-DNA circular
DEFINITION synthetic circular DNA
SOURCE synthetic DNA construct
ORGANISM synthetic DNA construct
REFERENCE 1 (bases 1 to 6722)
FEATURES Location/Qualifiers
source 1..6722
/organism="synthetic DNA construct"
/mol_type="other DNA"

enhancer 235..614
/label=CMV enhancer
/note="human cytomegalovirus immediate early enhancer"
promoter 615..818
/label=CMV promoter
/note="human cytomegalovirus (CMV) immediate early promoter"
promoter 863..881
/label=T7 promoter
/note="promoter for bacteriophage T7 RNA polymerase"
misc_feature 927..954
/label=Homology 1_pCDNA3.1
primer_bind 954..976
/label=ADARB1_lcv2_fw
primer_bind 955..976
/label=ADAR2CD-Cas9_HindIII_F
primer_bind 958..983
/label=AXC_lcv2_EFS-NS_fw
primer_bind 960..983
/label=Adar_out_forward_lv2
CDS 961..2100
/codon_start=1
/label=ADARB1(E488Q)_Catalytic Domain

/translation="MLADAVSRLVLGKFGDLTDNFSSPHARRKVLAGVVMTTGTDVKDA
KVISVSTGTKCINGEYMSDRGLALNDCHAEIISRRSLLRFLYTQLELYLNNKDDQKRSI
FQKSERGGFRLKENVQFHLYISTSPCGDARIFSPHEPILEEPADRHPNRKARGQLRTKI
ESGQGTIPVRSNASIQTWDGVLQGERLLTMSCSDKIARWNVVGIQGSLLSIFVEPIYFS
SIILGSLYHGDHLSRAMYQRISNIEDLPPLYTLNKPLLSGISNAEARQPGKAPNFSVNW
TVGDSAIEVINATTGKDELGRASRLCKHALYCRWMRVHGKVPSHLLRSKITKPNVYHES
KLAAKEYQAAKARLFTAFIKAGLGAWVEKPTEQDQFSLTP"
primer_bind 1324..1346
/label=E488Q_ADAR2_Mut_seq
primer_bind complement(1426..1447)
/label=E488Q_Mut_Classic_R
primer_bind 1440..1478
/label=E488Q_Mutagenesis_F
primer_bind complement(1440..1478)
/label=E488Q_Mutagenesis_R
primer_bind 1448..1472
/label=E488Q_Mut_Classic_F

CDS 2101..2148
/codon_start=1
/label=XTEN
/translation="SGSETPGTSESATPES"
primer_bind complement(2129..2148)
/label=ADAR2_CD_Inverse_R
primer_bind 2149..2170
/label=ADAR2_CD_Inverse_F
CDS 2152..2178
/codon_start=1
/product="HA (human influenza hemagglutinin) epitope tag"
/label=HA
/translation="YPYDVPDYA"
primer_bind complement(2170..2192)
/label=AXC_NLSout_NESin_R
primer_bind complement(2170..2192)
/label=Primer 1
CDS 2197..2217
/codon_start=1
/product="nuclear localization signal of SV40 large T
antigen"
/label=SV40 NLS
/translation="PKKKRKV"
CDS 2224..2244
/codon_start=1
/product="nuclear localization signal of SV40 large T
antigen"
/label=SV40 NLS
/translation="PKKKRKV"
primer_bind 2245..2267
/label=AXC_NLSout_NESin_F
misc_feature 2254..2288
/label=Homology 2_pCDNA3.1
polyA_signal 2322..2546
/label=bGH poly(A) signal
/note="bovine growth hormone polyadenylation signal"
rep_origin 2592..3020
/direction=RIGHT
/label=f1 ori
/note="f1 bacteriophage origin of replication; arrow
indicates direction of (+) strand synthesis"

promoter 3034..3363
/label=SV40 promoter
/note="SV40 enhancer and early promoter"
rep_origin 3214..3349
/label=SV40 ori
/note="SV40 origin of replication"
CDS 3430..4224
/codon_start=1
/gene="aph(3')-II (or nptII)"
/product="aminoglycoside phosphotransferase from Tn5"
/label=NeoR/KanR
/note="confers resistance to neomycin, kanamycin, and G418
(Geneticin(R))"

/translation="MIEQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQGRP
VLFVKTDLSGALNELQDEAARLSWLATTGVPCAAVLDVVTEAGRDWLLLGEVPGQDLLS
SHLAPAEKVSIMADAMRRLHTLDPATCPFDHQAKHRIERARTRMEAGLVDQDDLDEEHQ
GLAPAELFARLKARMPDGEDLVVTHGDACLPNIMVENGRFSGFIDCGRLGVADRYQDIA
LATRDIAEELGGEWADRFLVLYGIAAPDSQRIAFYRLLDEFF"
polyA_signal 4398..4519
/label=SV40 poly(A) signal
/note="SV40 polyadenylation signal"
primer_bind complement(4568..4584)
/label=M13 rev
/note="common sequencing primer, one of multiple similar
variants"
protein_bind 4592..4608
/label=lac operator
/bound_moiety="lac repressor encoded by lacI"
/note="The lac repressor binds to the lac operator to
inhibit transcription in E. coli. This inhibition can be
relieved by adding lactose or
isopropyl-beta-D-thiogalactopyranoside (IPTG)."
promoter complement(4616..4646)
/label=lac promoter
/note="promoter for the E. coli lac operon"
protein_bind 4661..4682
/label=CAP binding site
/bound_moiety="E. coli catabolite activator protein"
/note="CAP binding activates transcription in the presence
of cAMP."

rep_origin complement(4970..5555)
/direction=LEFT
/label=ori
/note="high-copy-number ColE1/pMB1/pBR322/pUC origin of replication"
CDS complement(5726..6586)
/codon_start=1
/gene="bla"
/product="beta-lactamase"
/label=AmpR
/note="confers resistance to ampicillin, carbenicillin, and related antibiotics"

/translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGYI
ELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVEYS
PVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRLDRW
EPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSA
LPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGAS
LIKHW"
promoter complement(6587..6691)
/gene="bla"
/label=AmpR promoter
ORIGIN
1 gacggatcgg gagatctccc gatcccctat ggtgcactct cagtacaatc tgctctgatg
61 ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgct gagtagtgcg
121 cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc
181 ttagggttag gcgttttgcg ctgcttcgcg atgtacgggc cagatatacg cgttgacatt
241 gattattgac tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata
301 tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc
361 cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc
421 attgacgtca atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt
481 atcatatgcc aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt
541 atgcccagta catgacctta tgggactttc ctacttggca gtacatctac gtattagtca
601 tcgctattac catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg
661 actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc
721 aaaatcaacg ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg
781 gtaggcgtgt acggtgggag gtctatataa gcagagctct ctggctaact agagaaccca
841 ctgcttactg gcttatcgaa attaatacga ctcactatag ggagacccaa gctggctagc
901 gtttaaacgg gccctctaga ctcgagcggc cgccactgtg ctggatatct gcagagaacc
961 atgttagctg acgctgtctc acgcctggtc ctgggtaagt ttggtgacct gaccgacaac
1021 ttctcctccc ctcacgctcg cagaaaagtg ctggctggag tcgtcatgac aacaggcaca

1081 gatgttaaag atgccaaggt gataagtgtt tctacaggaa caaaatgtat taatggtgaa
1141 tacatgagtg atcgtggcct tgcattaaat gactgccatg cagaaataat atctcggaga
1201 tccttgctca gatttcttta tacacaactt gagctttact taaataacaa agatgatcaa
1261 aaaagatcca tctttcagaa atcagagcga ggggggttta ggctgaagga gaatgtccag
1321 tttcatctgt acatcagcac ctctccctgt ggagatgcca gaatcttctc accacatgag
1381 ccaatcctgg aagaaccagc agatagacac ccaaatcgta aagcaagagg acagctacgg
1441 accaaaatag agtctggtca ggggacgatt ccagtgcgct ccaatgcgag catccaaacg
1501 tgggacgggg tgctgcaagg ggagcggctg ctcaccatgt cctgcagtga caagattgca
1561 cgctggaacg tggtgggcat ccagggatcc ctgctcagca ttttcgtgga gcccatttac
1621 ttctcgagca tcatcctggg cagcctttac cacggggacc acctttccag ggccatgtac
1681 cagcggatct ccaacataga ggacctgcca cctctctaca ccctcaacaa gcctttgctc
1741 agtggcatca gcaatgcaga agcacggcag ccagggaagg cccccaactt cagtgtcaac
1801 tggacggtag gcgactccgc tattgaggtc atcaacgcca cgactgggaa ggatgagctg
1861 ggccgcgcgt cccgcctgtg taagcacgcg ttgtactgtc gctggatgcg tgtgcacggc
1921 aaggttccct cccacttact acgctccaag attaccaagc ccaacgtgta ccatgagtcc
1981 aagctggcgg caaaggagta ccaggccgcc aaggcgcgtc tgttcacagc cttcatcaag
2041 gcggggctgg gggcctgggt ggagaagccc accgagcagg accagttctc actcacgccc
2101 agtggaagtg agacaccggg aacctcagag agcgccacgc cagaaagcgc ctatccctat
2161 gacgtgcccg attatgccag cctgggcagc ggctccccca agaaaaaacg caaggtggaa
2221 gatcctaaga aaaagcggaa agtggacgtg taaccaccac actggactag tggatccgag
2281 ctcggtacca agcttaagtt taaaccgctg atcagcctcg actgtgcctt ctagttgcca
2341 gccatctgtt gtttgcccct cccccgtgcc ttccttgacc ctggaaggtg ccactcccac
2401 tgtcctttcc taataaaatg aggaaattgc atcgcattgt ctgagtaggt gtcattctat
2461 tctgggggggt ggggtggggc aggacagcaa gggggaggat tgggaagaca atagcaggca
2521 tgctggggat gcggtgggct ctatggcttc tgaggcggaa agaaccagct ggggctctag
2581 ggggtatccc cacgcgccct gtagcggcgc attaagcgcg gcgggtgtgg tggttacgcg
2641 cagcgtgacc gctacacttg ccagcgccct agcgcccgct cctttcgctt tcttcccttc
2701 ctttctcgcc acgttcgccg gctttccccg tcaagctcta aatcgggggc tccctttagg
2761 gttccgattt agtgctttac ggcacctcga ccccaaaaaa cttgattagg gtgatggttc
2821 acgtagtggg ccatcgccct gatagacggt ttttcgccct ttgacgttgg agtccacgtt
2881 ctttaatagt ggactcttgt tccaaactgg aacaacactc aaccctatct cggtctattc
2941 ttttgattta taagggattt tgccgatttc ggcctattgg ttaaaaaatg agctgattta
3001 acaaaaattt aacgcgaatt aattctgtgg aatgtgtgtc agttagggtg tggaaagtcc
3061 ccaggctccc cagcaggcag aagtatgcaa agcatgcatc tcaattagtc agcaaccagg
3121 tgtggaaagt ccccaggctc cccagcaggc agaagtatgc aaagcatgca tctcaattag
3181 tcagcaacca tagtcccgcc cctaactccg cccatcccgc ccctaactcc gcccagttcc
3241 gcccattctc cgccccatgg ctgactaatt ttttttattt atgcagaggc cgaggccgcc
3301 tctgcctctg agctattcca gaagtagtga ggaggctttt ttggaggcct aggcttttgc
3361 aaaaagctcc cgggagcttg tatatccatt ttcggatctg atcaagagac aggatgagga
3421 tcgtttcgca tgattgaaca agatggattg cacgcaggtt ctccggccgc ttgggtggag
3481 aggctattcg gctatgactg ggcacaacag acaatcggct gctctgatgc cgccgtgttc

3541 cggctgtcag cgcaggggcg cccggttctt tttgtcaaga ccgacctgtc cggtgccctg
3601 aatgaactgc aggacgaggc agcgcggcta tcgtggctgg ccacgacggg cgttccttgc
3661 gcagctgtgc tcgacgttgt cactgaagcg ggaagggact ggctgctatt gggcgaagtg
3721 ccggggcagg atctcctgtc atctcacctt gctcctgccg agaaagtatc catcatggct
3781 gatgcaatgc ggcggctgca tacgcttgat ccggctacct gcccattcga ccaccaagcg
3841 aaacatcgca tcgagcgagc acgtactcgg atggaagccg gtcttgtcga tcaggatgat
3901 ctggacgaag agcatcaggg gctcgcgcca gccgaactgt tcgccaggct caaggcgcgc
3961 atgcccgacg gcgaggatct cgtcgtgacc catggcgatg cctgcttgcc gaatatcatg
4021 gtggaaaatg gccgcttttc tggattcatc gactgtggcc ggctgggtgt ggcggaccgc
4081 tatcaggaca tagcgttggc tacccgtgat attgctgaag agcttggcgg cgaatgggct
4141 gaccgcttcc tcgtgcttta cggtatcgcc gctcccgatt cgcagcgcat cgccttctat
4201 cgccttcttg acgagttctt ctgagcggga ctctggggtt cgaaatgacc gaccaagcga
4261 cgcccaacct gccatcacga gatttcgatt ccaccgccgc cttctatgaa aggttgggct
4321 tcggaatcgt tttccgggac gccggctgga tgatcctcca gcgcggggat ctcatgctgg
4381 agttcttcgc ccaccccaac ttgtttattg cagcttataa tggttacaaa taaagcaata
4441 gcatcacaaa tttcacaaat aaagcatttt tttcactgca ttctagttgt ggtttgtcca
4501 aactcatcaa tgtatcttat catgtctgta taccgtcgac ctctagctag agcttggcgt
4561 aatcatggtc atagctgttt cctgtgtgaa attgttatcc gctcacaatt ccacacaaca
4621 tacgagccgg aagcataaag tgtaaagcct ggggtgccta atgagtgagc taactcacat
4681 taattgcgtt gcgctcactg cccgctttcc agtcgggaaa cctgtcgtgc cagctgcatt
4741 aatgaatcgg ccaacgcgcg gggagaggcg gtttgcgtat tgggcgctct tccgcttcct
4801 cgctcactga ctcgctgcgc tcggtcgttc ggctgcggcg agcggtatca gctcactcaa
4861 aggcggtaat acggttatcc acagaatcag gggataacgc aggaaagaac atgtgagcaa
4921 aaggccagca aaaggccagg aaccgtaaaa aggccgcgtt gctggcgttt ttccataggc
4981 tccgcccccc tgacgagcat cacaaaaatc gacgctcaag tcagaggtgg cgaaacccga
5041 caggactata aagataccag gcgtttcccc ctggaagctc cctcgtgcgc tctcctgttc
5101 cgaccctgcc gcttaccgga tacctgtccg cctttctccc ttcgggaagc gtggcgcttt
5161 ctcatagctc acgctgtagg tatctcagtt cggtgtaggt cgttcgctcc aagctgggct
5221 gtgtgcacga acccccccgtt cagcccgacc gctgcgcctt atccggtaac tatcgtcttg
5281 agtccaaccc ggtaagacac gacttatcgc cactggcagc agccactggt aacaggatta
5341 gcagagcgag gtatgtaggc ggtgctacag agttcttgaa gtggtggcct aactacggct
5401 acactagaag aacagtattt ggtatctgcg ctctgctgaa gccagttacc ttcggaaaaa
5461 gagttggtag ctcttgatcc ggcaaacaaa ccaccgctgg tagcggtttt tttgtttgca
5521 agcagcagat tacgcgcaga aaaaaaggat ctcaagaaga tcctttgatc ttttctacgg
5581 ggtctgacgc tcagtggaac gaaaactcac gttaagggat tttggtcatg agattatcaa
5641 aaaggatctt cacctagatc cttttaaatt aaaaatgaag ttttaaatca atctaaagta
5701 tatatgagta aacttggtct gacagttacc aatgcttaat cagtgaggca cctatctcag
5761 cgatctgtct atttcgttca tccatagttg cctgactccc cgtcgtgtag ataactacga
5821 tacgggaggg cttaccatct ggccccagtg ctgcaatgat accgcgagac ccacgctcac
5881 cggctccaga tttatcagca ataaaccagc cagccggaag ggccgagcgc agaagtggtc
5941 ctgcaacttt atccgcctcc atccagtcta ttaattgttg ccgggaagct agagtaagta

6001 gttcgccagt taatagtttg cgcaacgttg ttgccattgc tacaggcatc gtggtgtcac
6061 gctcgtcgtt tggtatggct tcattcagct ccggttccca acgatcaagg cgagttacat
6121 gatcccccat gttgtgcaaa aaagcggtta gctccttcgg tcctccgatc gttgtcagaa
6181 gtaagttggc cgcagtgtta tcactcatgg ttatggcagc actgcataat tctcttactg
6241 tcatgccatc cgtaagatgc ttttctgtga ctggtgagta ctcaaccaag tcattctgag
6301 aatagtgtat gcggcgaccg agttgctctt gcccggcgtc aatacgggat aataccgcgc
6361 cacatagcag aactttaaaa gtgctcatca ttggaaaacg ttcttcgggg cgaaaactct
6421 caaggatctt accgctgttg agatccagtt cgatgtaacc cactcgtgca cccaactgat
6481 cttcagcatc ttttactttc accagcgttt ctgggtgagc aaaaacagga aggcaaaatg
6541 ccgcaaaaaa gggaataagg gcgacacgga aatgttgaat actcatactc ttccttttc
6601 aatattattg aagcatttat cagggttatt gtctcatgag cggatacata tttgaatgta
6661 tttagaaaaa taaacaaata ggggttccgc gcacatttcc ccgaaaagtg ccacctgacg
6721 tc

50bp_GFP_mCherry_extension (SEQ ID NO: 31).

LOCUS Exported 4951 bp ds-DNA circular
DEFINITION synthetic circular DNA
SOURCE synthetic DNA construct
ORGANISM recombinant plasmid
REFERENCE 1 (bases 1 to 4951)
FEATURES Location/Qualifiers
source 1..4951
/organism="recombinant plasmid"
/mol_type="other DNA"
primer_bind 1..40
/label=EF1a_Gibson_F
primer_bind 1..20
/label=Primer 2
misc_feature 1..7
/label=sgRNA scaffold_termination
promoter 21..566
/label=EF1a promoter
primer_bind complement(554..591)
/label=EF1a_Gibson_R
CDS 572..1282
/codon_start=1
/product="monomeric derivative of DsRed fluorescent protein
(Shaner et al., 2004)"

/label=mCherry
/note="mammalian codon-optimized"

/translation="MVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEG
TQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNF
EDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALK
GEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERA
EGRHSTGGMDELYK"

primer_bind 572..591
/label=mCherry_BGH_F
primer_bind complement(1259..1306)
/label=Primer 1
primer_bind complement(1259..1286)
/label=mCherry_P2A_Gib_R
primer_bind complement(1259..1282)
/label=mCherry_HindIII_R
misc_feature 1283..1306
/label=Gibson Overlap
primer_bind 1283..1301
/label=mCherry_P2A_Gib_F
polyA_signal 1330..1554
/label=bGH poly(A) signal
/note="bovine growth hormone polyadenylation signal"
primer_bind complement(1535..1573)
/label=mCherry_BGH_Gib_R
primer_bind complement(1535..1554)
/label=mCherry_BGH_R
primer_bind complement(1536..1555)
/label=bGH_NotI_R
primer_bind complement(1558..1573)
/label=SK primer
/note="common sequencing primer, one of multiple similar
variants"
primer_bind complement(1608..1627)
/label=T3
primer_bind complement(1645..1665)
/label=M13-rev
misc_binding complement(1671..1693)
/label=LacO
promoter complement(1698..1727)
/label=lac

rep_origin complement(2033..2661)
/direction=LEFT
/label=ColE1 origin
CDS complement(2813..3472)
/label=AmpR
promoter complement(3712..3740)
/label=Amp prom
rep_origin 3811..4251
/direction=RIGHT
/label=F1 ori
CDS complement(4258..4326)
/label=LacZ alpha
primer_bind 4397..4414
/label=M13-fwd
primer_bind 4424..4443
/label=T7
promoter 4555..4817
/label=U6 promoter
primer_bind 4798..4864
/label=no_spacer_universal_scaff_f
primer_bind 4803..4862
/label=50bp_GFP_F
primer_bind 4803..4862
/label=50bp_GFP_revcomp_F(+G)
primer_bind 4803..4862
/label=10bp_GFP_spacer_F
primer_bind 4803..4862
/label=30bp_GFP_spacer_F
primer_bind 4803..4862
/label=70bp_GFP_spacer_F
primer_bind 4803..4862
/label=ACTB_3_ext_CgRNA_For
primer_bind complement(4803..4817)
/label=Primer 3
primer_bind complement(4803..4817)
/label=extension_gibson_R
misc_feature 4818..4838
/label=50bp_EGFP_targeting_spacer
misc_feature 4839..4924
/label=sgRNA scaffold
primer_bind 4839..4865

/label=scaffold_3_ext_template_For
primer_bind complement(4912..4930)
/label=scaffold_3_ext_template_Rev
primer_bind complement(join(4913..4951,1..20))
/label=eGFP_3_ext_R
primer_bind complement(join(4913..4951,1..20))
/label=gfp_3_extension_revcomp
primer_bind complement(join(4913..4951,1..20))
/label=ACTB_3_ext_AgRNA_Rev
misc_feature 4925..4930
/label=Linker
misc_feature 4931..4951
/label=EGFP_extension
ORIGIN
1 ttttttcct gcagcccggg aaggatctgc gatcgctccg gtgcccgtca gtgggcagag
61 cgcacatcgc ccacagtccc cgagaagttg gggggagggg tcggcaattg aacgggtgcc
121 tagagaaggt ggcgcggggt aaactgggaa agtgatgtcg tgtactggct ccgccttttt
181 cccgagggtg ggggagaacc gtatataagt gcagtagtcg ccgtgaacgt tcttttcgc
241 aacgggtttg ccgccagaac acagctgaag cttcgagggg ctcgcatctc tccttcacgc
301 gcccgccgcc ctacctgagg ccgccatcca cgccggttga gtcgcgttct gccgcctccc
361 gcctgtggtg cctcctgaac tgcgtccgcc gtctaggtaa gtttaaagct caggtcgaga
421 ccgggccttt gtccggcgct cccttggagc ctacctagac tcagccggct ctccacgctt
481 tgcctgaccc tgcttgctca actctacgtc tttgtttcgt tttctgttct gcgccgttac
541 agatccaagc tgtgaccggc gcctacgcta gatggtgagc aagggcgagg aggataacat
601 ggccatcatc aaggagttca tgcgcttcaa ggtgcacatg gagggctccg tgaacggcca
661 cgagttcgag atcgagggcg agggcgaggg ccgcccctac gagggcaccc agaccgccaa
721 gctgaaggtg accaagggtg gcccccctgcc cttcgcctgg gacatcctgt cccctcagtt
781 catgtacggc tccaaggcct acgtgaagca ccccgccgac atccccgact acttgaagct
841 gtccttcccc gagggcttca agtgggagcg cgtgatgaac ttcgaggacg gcggcgtggt
901 gaccgtgacc caggactcct ccctgcagga cggcgagttc atctacaagg tgaagctgcg
961 cggcaccaac ttcccctccg acggccccgt aatgcagaag aagaccatgg gctgggaggc
1021 ctcctccgag cggatgtacc ccgaggacgg cgccctgaag ggcgagatca agcagaggct
1081 gaagctgaag gacggcggcc actacgacgc tgaggtcaag accacctaca aggccaagaa
1141 gcccgtgcag ctgcccggcg cctacaacgt caacatcaag ttggacatca cctcccacaa
1201 cgaggactac accatcgtgg aacagtacga acgcgccgag ggccgccact ccaccggcgg
1261 catggacgag ctgtacaagt aatccgagct cggtaccaag cttaagttta aaccgctgat
1321 cagcctcgac tgtgccttct agttgccagc catctgttgt ttgcccctcc cccgtgcctt
1381 ccttgaccct ggaaggtgcc actcccactg tcctttccta ataaaatgag gaaattgcat
1441 cgcattgtct gagtaggtgt cattctattc tgggggggtgg ggtggggcag gacagcaagg
1501 gggaggattg ggaagacaat agcaggcatg ctggggatgc ggtgggctct atgggggatc
1561 cactagttct agagcggccg ccaccgcggt ggagctccag cttttgttcc ctttagtgag

1621 ggttaattgc gcgcttggcg taatcatggt catagctgtt tcctgtgtga aattgttatc
1681 cgctcacaat tccacacaac atacgagccg gaagcataaa gtgtaaagcc tggggtgcct
1741 aatgagtgag ctaactcaca ttaattgcgt tgcgctcact gcccgctttc cagtcgggaa
1801 acctgtcgtg ccagctgcat taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta
1861 ttgggcgctc ttccgcttcc tcgctcactg actcgctgcg ctcggtcgtt cggctgcggc
1921 gagcggtatc agctcactca aaggcggtaa tacggttatc cacagaatca ggggataacg
1981 caggaaagaa catgtgagca aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt
2041 tgctggcgtt tttccatagg ctccgccccc ctgacgagca tcacaaaaat cgacgctcaa
2101 gtcagaggtg gcgaaacccg acaggactat aaagatacca ggcgtttccc cctggaagct
2161 ccctcgtgcg ctctcctgtt ccgaccctgc cgcttaccgg atacctgtcc gcctttctcc
2221 cttcgggaag cgtggcgctt tctcatagct cacgctgtag gtatctcagt tcggtgtagg
2281 tcgttcgctc caagctgggc tgtgtgcacg aacccccccgt tcagcccgac cgctgcgcct
2341 tatccggtaa ctatcgtctt gagtccaacc cggtaagaca cgacttatcg ccactggcag
2401 cagccactgg taacaggatt agcagagcga ggtatgtagg cggtgctaca gagttcttga
2461 agtggtggcc taactacggc tacactagaa ggacagtatt tggtatctgc gctctgctga
2521 agccagttac cttcggaaaa agagttggta gctcttgatc cggcaaacaa accaccgctg
2581 gtagcggtgg ttttttttgtt tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag
2641 aagatccttt gatcttttct acggggtctg acgctcagtg gaacgaaaac tcacgttaag
2701 ggattttggt catgagatta tcaaaaagga tcttcaccta gatcctttta aattaaaaat
2761 gaagttttaa atcaatctaa agtatatatg agtaaacttg gtctgacagt taccaatgct
2821 taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcatccata gttgcctgac
2881 tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa
2941 tgataccgcg agacccacgc tcaccggctc cagatttatc agcaataaac cagccagccg
3001 gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt
3061 gttgccggga agctagagta agtagttcgc cagttaatag tttgcgcaac gttgttgcca
3121 ttgctacagg catcgtggtg tcacgctcgt cgtttggtat ggcttcattc agctccggtt
3181 cccaacgatc aaggcgagtt acatgatccc ccatgttgtg caaaaaagcg gttagctcct
3241 tcggtcctcc gatcgttgtc agaagtaagt tggccgcagt gttatcactc atggttatgg
3301 cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct gtgactggtg
3361 agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg
3421 cgtcaatacg ggataatacc gcgccacata gcagaacttt aaaagtgctc atcattggaa
3481 aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt
3541 aacccactcg tgcacccaac tgatcttcag catcttttac tttcaccagc gtttctgggt
3601 gagcaaaaac aggaaggcaa aatgccgcaa aaaagggaat aagggcgaca cggaaatgtt
3661 gaatactcat actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca
3721 tgagcggata catatttgaa tgtatttaga aaaataaaca aatagggttt ccgcgcacat
3781 ttccccgaaa agtgccacct aaattgtaag cgttaatatt ttgttaaaat tcgcgttaaa
3841 tttttgttaa atcagctcat tttttaacca ataggccgaa atcggcaaaa tcccttataa
3901 atcaaaagaa tagaccgaga tagggttgag tgttgttcca gtttggaaca agagtccact
3961 attaaagaac gtggactcca acgtcaaagg gcgaaaaacc gtctatcagg gcgatggccc
4021 actacgtgaa ccatcaccct aatcaagttt tttggggtcg aggtgccgta aagcactaaa

4081 tcggaaccct aaagggagcc cccgatttag agcttgacgg ggaaagccgg cgaacgtggc
4141 gagaaaggaa gggaagaaag cgaaaggagc gggcgctagg gcgctggcaa gtgtagcggt
4201 cacgctgcgc gtaaccacca cacccgccgc gcttaatgcg ccgctacagg gcgcgtccca
4261 ttcgccattc aggctgcgca actgttggga agggcgatcg gtgcgggcct cttcgctatt
4321 acgccagctg gcgaaagggg gatgtgctgc aaggcgatta agttgggtaa cgccagggtt
4381 ttcccagtca cgacgttgta aaacgacggc cagtgagcgc gcgtaatacg actcactata
4441 gggcgaattg ggtaccgggc ccccctcga ggtcgacggt atcgataagc ttgatatcgt
4501 gtacaaaaaa gcaggcttta aaggaaccaa ttcagtcgac tggatccggt accaaggtcg
4561 ggcaggaaga gggcctattt cccatgattc cttcatattt gcatatacga tacaaggctg
4621 ttagagagat aattagaatt aatttgactg taaacacaaa gatattagta caaaatacgt
4681 gacgtagaaa gtaataattt cttgggtagt ttgcagtttt aaaattatgt tttaaaatgg
4741 actatcatat gcttaccgta acttgaaagt atttcgattt cttggcttta tatatcttgt
4801 ggaaaggacg aaacaccgaa gtcatgccgt ttcatgtggt ttaagagcta tgctggaaac
4861 agcatagcaa gtttaaataa ggctagtccg ttatcaactt gaaaaagtgg caccgagtcg
4921 gtgcttcatt gtgtcggcca cggaacaggc a

spacerless_GFP_mCherry_extension (SEQ ID NO: 32).

LOCUS Exported 4930 bp ds-DNA circular
DEFINITION synthetic circular DNA
SOURCE synthetic DNA construct
ORGANISM recombinant plasmid
REFERENCE 1 (bases 1 to 4930)
FEATURES Location/Qualifiers
source 1..4930
/organism="recombinant plasmid"
/mol_type="other DNA"
rep_origin 13..453
/direction=RIGHT
/label=F1 ori
CDS complement(460..528)
/label=LacZ alpha
primer_bind 599..616
/label=M13-fwd
primer_bind 626..645
/label=T7
promoter 757..1019
/label=U6 promoter
primer_bind complement(998..1019)
/label=scaffold_out_R
primer_bind 1000..1045

/label=no_spacer_universal_scaff_f
primer_bind 1005..1043
/label=50bp_GFP_F
primer_bind 1005..1043
/label=ACTB_3_ext_CgRNA_For
misc_feature 1020..1105
/label=sgRNA scaffold
primer_bind 1020..1046
/label=scaffold_3_ext_template_For
primer_bind complement(1093..1111)
/label=scaffold_3_ext_template_Rev
primer_bind complement(1094..1152)
/label=eGFP_3_ext_R
primer_bind complement(1094..1152)
/label=gfp_3_extension_revcomp
primer_bind complement(1094..1152)
/label=ACTB_3_ext_AgRNA_Rev
misc_feature 1106..1111
/label=Linker
misc_feature 1112..1132
/label=EGFP_extension
primer_bind 1133..1172
/label=EF1a_Gibson_F
primer_bind 1133..1152
/label=3_ext_backbone_For
misc_feature 1133..1139
/label=sgRNA scaffold termination
promoter 1153..1698
/label=EF1a promoter
primer_bind complement(1686..1723)
/label=EF1a_Gibson_R
CDS 1704..2414
/codon_start=1
/product="monomeric derivative of DsRed fluorescent protein
(Shaner et al., 2004)"
/label=mCherry
/note="mammalian codon-optimized"

/translation="MVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEG
TQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNF
EDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALK

GEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERA
EGRHSTGGMDELYK"
primer_bind 1704..1723
/label=mCherry_BGH_F
primer_bind complement(2391..2438)
/label=Primer 1
primer_bind complement(2391..2414)
/label=mCherry_HindIII_R
misc_feature 2415..2438
/label=Gibson Overlap
polyA_signal 2462..2686
/label=bGH poly(A) signal
/note="bovine growth hormone polyadenylation signal"
primer_bind complement(2667..2705)
/label=mCherry_BGH_Gib_R
primer_bind complement(2667..2686)
/label=mCherry_BGH_R
primer_bind complement(2668..2687)
/label=bGH_NotI_R
primer_bind complement(2690..2705)
/label=SK primer
/note="common sequencing primer, one of multiple similar
variants"
primer_bind complement(2740..2759)
/label=T3
primer_bind complement(2777..2797)
/label=M13-rev
misc_binding complement(2803..2825)
/label=LacO
promoter complement(2830..2859)
/label=lac
rep_origin complement(3165..3793)
/direction=LEFT
/label=ColE1 origin
CDS complement(3945..4604)
/label=AmpR
promoter complement(4844..4872)
/label=Amp prom
ORIGIN
1 ctaaattgta agcgttaata ttttgttaaa attcgcgtta aatttttgtt aaatcagctc
61 atttttaac caataggccg aaatcggcaa aatcccttat aaatcaaaag aatagaccga

121 gatagggttg agtgttgttc cagtttggaa caagagtcca ctattaaaga acgtggactc
181 caacgtcaaa gggcgaaaaa ccgtctatca gggcgatggc ccactacgtg aaccatcacc
241 ctaatcaagt ttttggggt cgaggtgccg taaagcacta aatcggaacc ctaaagggag
301 cccccgattt agagcttgac ggggaaagcc ggcgaacgtg gcgagaaagg aagggaagaa
361 agcgaaagga gcgggcgcta gggcgctggc aagtgtagcg gtcacgctgc gcgtaaccac
421 cacacccgcc gcgcttaatg cgccgctaca gggcgcgtcc cattcgccat tcaggctgcg
481 caactgttgg gaagggcgat cggtgcgggc ctcttcgcta ttacgccagc tggcgaaagg
541 gggatgtgct gcaaggcgat taagttgggt aacgccaggg ttttcccagt cacgacgttg
601 taaaacgacg gccagtgagc gcgcgtaata cgactcacta tagggcgaat tgggtaccgg
661 gccccccctc gaggtcgacg gtatcgataa gcttgatatc gtgtacaaaa aagcaggctt
721 taaaggaacc aattcagtcg actggatccg gtaccaaggt cgggcaggaa gagggcctat
781 ttcccatgat tccttcatat ttgcatatac gatacaaggc tgttagagag ataattagaa
841 ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga aagtaataat
901 ttcttgggta gtttgcagtt ttaaaattat gttttaaaat ggactatcat atgcttaccg
961 taacttgaaa gtatttcgat ttcttggctt tatatatctt gtggaaagga cgaaacaccg
1021 tttaagagct atgctggaaa cagcatagca agtttaaata aggctagtcc gttatcaact
1081 tgaaaaagtg gcaccgagtc ggtgcttcat tgtgtcggcc acggaacagg catttttttc
1141 ctgcagcccg ggaaggatct gcgatcgctc cggtgcccgt cagtgggcag agcgcacatc
1201 gcccacagtc cccgagaagt tgggggggagg ggtcggcaat tgaacgggtg cctagagaag
1261 gtggcgcggg gtaaactggg aaagtgatgt cgtgtactgg ctccgccttt ttcccgaggg
1321 tgggggagaa ccgtatataa gtgcagtagt cgccgtgaac gttcttttc gcaacgggtt
1381 tgccgccaga acacagctga agcttcgagg ggctcgcatc tctccttcac gcgcccgccg
1441 ccctacctga ggccgccatc cacgccggtt gagtcgcgtt ctgccgcctc ccgcctgtgg
1501 tgcctcctga actgcgtccg ccgtctaggt aagtttaaag ctcaggtcga gaccgggcct
1561 ttgtccggcg ctcccttgga gcctacctag actcagccgg ctctccacgc tttgcctgac
1621 cctgcttgct caactctacg tctttgtttc gttttctgtt ctgcgccgtt acagatccaa
1681 gctgtgaccg gcgcctacgc tagatggtga gcaagggcga ggaggataac atggccatca
1741 tcaaggagtt catgcgcttc aaggtgcaca tggagggctc cgtgaacggc cacgagttcg
1801 agatcgaggg cgagggcgag ggccgccccc acgagggcac ccagaccgcc aagctgaagg
1861 tgaccaaggg tggccccctg cccttcgcct gggacatcct gtcccctcag ttcatgtacg
1921 gctccaaggc ctacgtgaag caccccgccg acatccccga ctacttgaag ctgtccttcc
1981 ccgagggctt caagtgggag cgcgtgatga acttcgagga cggcggcgtg gtgaccgtga
2041 cccaggactc ctccctgcag gacggcgagt tcatctacaa ggtgaagctg cgcggcacca
2101 acttcccctc cgacggcccc gtaatgcaga agaagaccat gggctgggag gcctcctccg
2161 agcggatgta ccccgaggac ggcgccctga agggcgagat caagcagagg ctgaagctga
2221 aggacggcgg ccactacgac gctgaggtca agaccaccta caaggccaag aagcccgtgc
2281 agctgcccgg cgcctacaac gtcaacatca agttggacat cacctcccac aacgaggact
2341 acaccatcgt ggaacagtac gaacgcgccg agggccgcca ctccaccggc ggcatggacg
2401 agctgtacaa gtaatccgag ctcggtacca agcttaagtt taaaccgctg atcagcctcg
2461 actgtgcctt ctagttgcca gccatctgtt gtttgcccct cccccgtgcc ttccttgacc
2521 ctggaaggtg ccactcccac tgtcctttcc taataaaatg aggaaattgc atcgcattgt

2581 ctgagtaggt gtcattctat tctggggggt ggggtggggc aggacagcaa gggggaggat
2641 tgggaagaca atagcaggca tgctggggat gcggtgggct ctatgggggа tccactagtt
2701 ctagagcggc cgccaccgcg gtggagctcc agcttttgtt ccctttagtg agggttaatt
2761 gcgcgcttgg cgtaatcatg gtcatagctg tttcctgtgt gaaattgtta tccgctcaca
2821 attccacaca acatacgagc cggaagcata aagtgtaaag cctggggtgc ctaatgagtg
2881 agctaactca cattaattgc gttgcgctca ctgcccgctt tccagtcggg aaacctgtcg
2941 tgccagctgc attaatgaat cggccaacgc gcggggagag gcggtttgcg tattgggcgc
3001 tcttccgctt cctcgctcac tgactcgctg cgctcggtcg ttcggctgcg gcgagcggta
3061 tcagctcact caaaggcggt aatacggtta tccacagaat caggggataa cgcaggaaag
3121 aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta aaaaggccgc gttgctggcg
3181 tttttccata ggctccgccc ccctgacgag catcacaaaa atcgacgctc aagtcagagg
3241 tggcgaaacc cgacaggact ataaagatac caggcgtttc cccctggaag ctccctcgtg
3301 cgctctcctg ttccgaccct gccgcttacc ggatacctgt ccgcctttct cccttcggga
3361 agcgtggcgc tttctcatag ctcacgctgt aggtatctca gttcggtgta ggtcgttcgc
3421 tccaagctgg gctgtgtgca cgaacccccc gttcagcccg accgctgcgc cttatccggt
3481 aactatcgtc ttgagtccaa cccggtaaga cacgacttat cgccactggc agcagccact
3541 ggtaacagga ttagcagagc gaggtatgta ggcggtgcta cagagttctt gaagtggtgg
3601 cctaactacg gctacactag aaggacagta tttggtatct gcgctctgct gaagccagtt
3661 accttcggaa aaagagttgg tagctcttga tccggcaaac aaaccaccgc tggtagcggt
3721 ggtttttttg tttgcaagca gcagattacg cgcagaaaaa aaggatctca agaagatcct
3781 ttgatctttt ctacggggtc tgacgctcag tggaacgaaa actcacgtta agggattttg
3841 gtcatgagat tatcaaaaag gatcttcacc tagatccttt taaattaaaa atgaagtttt
3901 aaatcaatct aaagtatata tgagtaaact tggtctgaca gttaccaatg cttaatcagt
3961 gaggcaccta tctcagcgat ctgtctattt cgttcatcca tagttgcctg actccccgtc
4021 gtgtagataa ctacgatacg ggagggctta ccatctggcc ccagtgctgc aatgataccg
4081 cgagacccac gctcaccggc tccagattta tcagcaataa accagccagc cggaagggcc
4141 gagcgcagaa gtggtcctgc aactttatcc gcctccatcc agtctattaa ttgttgccgg
4201 gaagctagag taagtagttc gccagttaat agtttgcgca acgttgttgc cattgctaca
4261 ggcatcgtgg tgtcacgctc gtcgtttggt atggcttcat tcagctccgg ttcccaacga
4321 tcaaggcgag ttacatgatc ccccatgttg tgcaaaaaag cggttagctc cttcggtcct
4381 ccgatcgttg tcagaagtaa gttggccgca gtgttatcac tcatggttat ggcagcactg
4441 cataattctc ttactgtcat gccatccgta agatgctttt ctgtgactgg tgagtactca
4501 accaagtcat tctgagaata gtgtatgcgg cgaccgagtt gctcttgccc ggcgtcaata
4561 cgggataata ccgcgccaca tagcagaact ttaaaagtgc tcatcattgg aaaacgttct
4621 tcggggcgaa aactctcaag gatcttaccg ctgttgagat ccagttcgat gtaacccact
4681 cgtgcaccca actgatcttc agcatctttt actttcacca gcgtttctgg gtgagcaaaa
4741 acaggaaggc aaaatgccgc aaaaaaggga ataagggcga cacggaaatg ttgaatactc
4801 atactcttcc tttttcaata ttattgaagc atttatcagg gttattgtct catgagcgga
4861 tacatatttg aatgtattta gaaaaataaa caaatagggg ttccgcgcac atttccccga
4921 aaagtgccac

GFP_no_spacer_revcomp_mCherry_gibson (SEQ ID NO: 33).

LOCUS Exported 4930 bp ds-DNA circular
DEFINITION synthetic circular DNA
SOURCE synthetic DNA construct
ORGANISM recombinant plasmid
REFERENCE 1 (bases 1 to 4930)
FEATURES Location/Qualifiers
source 1..4930
/organism="recombinant plasmid"
/mol_type="other DNA"
primer_bind 1..20
/label=Primer 2
misc_feature 1..7
/label=sgRNA scaffold termination
promoter 21..566
/label=EF1a promoter
primer_bind complement(554..591)
/label=EF1a_Gibson_R
CDS 572..1282
/codon_start=1
/product="monomeric derivative of DsRed fluorescent protein
(Shaner et al., 2004)"
/label=mCherry
/note="mammalian codon-optimized"

/translation="MVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEG
TQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNF
EDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALK
GEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERA
EGRHSTGGMDELYK"
primer_bind 572..591
/label=mCherry_BGH_F
primer_bind complement(1259..1306)
/label=Primer 1
primer_bind complement(1259..1282)
/label=mCherry_HindIII_R
misc_feature 1283..1306
/label=Gibson Overlap
polyA_signal 1330..1554
/label=bGH poly(A) signal
/note="bovine growth hormone polyadenylation signal"

primer_bind complement(1535..1573)
/label=mCherry_BGH_Gib_R
primer_bind complement(1535..1554)
/label=mCherry_BGH_R
primer_bind complement(1536..1555)
/label=bGH_NotI_R
primer_bind complement(1558..1573)
/label=SK primer
/note="common sequencing primer, one of multiple similar variants"
primer_bind complement(1608..1627)
/label=T3
primer_bind complement(1645..1665)
/label=M13-rev
misc_binding complement(1671..1693)
/label=LacO
promoter complement(1698..1727)
/label=lac
rep_origin complement(2033..2661)
/direction=LEFT
/label=ColE1 origin
CDS complement(2813..3472)
/label=AmpR
promoter complement(3712..3740)
/label=Amp prom
rep_origin 3811..4251
/direction=RIGHT
/label=F1 ori
CDS complement(4258..4326)
/label=LacZ alpha
primer_bind 4397..4414
/label=M13-fwd
primer_bind 4424..4443
/label=T7
promoter 4555..4817
/label=U6 promoter
primer_bind 4798..4843
/label=no_spacer_universal_scaff_f
primer_bind 4803..4841
/label=ACTB_3_ext_CgRNA_For
primer_bind complement(4803..4817)

/label=Primer 3
primer_bind complement(4803..4817)
/label=extension_gibson_R
misc_feature 4818..4903
/label=sgRNA scaffold
primer_bind 4818..4844
/label=scaffold_3_ext_template_For
primer_bind complement(4891..4909)
/label=scaffold_3_ext_template_Rev
primer_bind complement(join(4892..4930,1..20))
/label=gfp_3_extension_revcomp
primer_bind complement(join(4892..4930,1..20))
/label=ACTB_3_ext_AgRNA_Rev
misc_feature 4904..4909
/label=Linker
misc_feature 4910..4930
/label=EGFP_revcomp_extension
primer_bind join(4930,1..40)
/label=EF1a_Gibson_F
ORIGIN
1 tttttttcct gcagcccggg aaggatctgc gatcgctccg gtgcccgtca gtgggcagag
61 cgcacatcgc ccacagtccc cgagaagttg gggggagggg tcggcaattg aacgggtgcc
121 tagagaaggt ggcgcggggt aaactgggaa agtgatgtcg tgtactggct ccgccttttt
181 cccgagggtg ggggagaacc gtatataagt gcagtagtcg ccgtgaacgt tcttttcgc
241 aacgggtttg ccgccagaac acagctgaag cttcgagggg ctcgcatctc tccttcacgc
301 gcccgccgcc ctacctgagg ccgccatcca cgccggttga gtcgcgttct gccgcctccc
361 gcctgtggtg cctcctgaac tgcgtccgcc gtctaggtaa gtttaaagct caggtcgaga
421 ccgggccttt gtccggcgct cccttggagc ctacctagac tcagccggct ctccacgctt
481 tgcctgaccc tgcttgctca actctacgtc tttgtttcgt tttctgttct gcgccgttac
541 agatccaagc tgtgaccggc gcctacgcta gatggtgagc aagggcgagg aggataacat
601 ggccatcatc aaggagttca tgcgcttcaa ggtgcacatg gagggctccg tgaacggcca
661 cgagttcgag atcgagggcg agggcgaggg ccgcccctac gagggcaccc agaccgccaa
721 gctgaaggtg accaagggtg gcccccctgcc cttcgcctgg gacatcctgt cccctcagtt
781 catgtacggc tccaaggcct acgtgaagca ccccgccgac atccccgact acttgaagct
841 gtccttcccc gagggcttca agtgggagcg cgtgatgaac ttcgaggacg gcggcgtggt
901 gaccgtgacc caggactcct ccctgcagga cggcgagttc atctacaagg tgaagctgcg
961 cggcaccaac ttcccctccg acggccccgt aatgcagaag aagaccatgg gctgggaggc
1021 ctcctccgag cggatgtacc ccgaggacgg cgccctgaag ggcgagatca agcagaggct
1081 gaagctgaag gacggcggcc actacgacgc tgaggtcaag accacctaca aggccaagaa
1141 gcccgtgcag ctgcccggcg cctacaacgt caacatcaag ttggacatca cctcccacaa
1201 cgaggactac accatcgtgg aacagtacga acgcgccgag ggccgccact ccaccggcgg

1261 catggacgag ctgtacaagt aatccgagct cggtaccaag cttaagttta aaccgctgat
1321 cagcctcgac tgtgccttct agttgccagc catctgttgt ttgcccctcc cccgtgcctt
1381 ccttgaccct ggaaggtgcc actcccactg tcctttccta ataaaatgag gaaattgcat
1441 cgcattgtct gagtaggtgt cattctattc tgggggggtgg ggtggggcag gacagcaagg
1501 gggaggattg ggaagacaat agcaggcatg ctggggatgc ggtgggctct atgggggatc
1561 cactagttct agagcggccg ccaccgcggt ggagctccag cttttgttcc ctttagtgag
1621 ggttaattgc gcgcttggcg taatcatggt catagctgtt tcctgtgtga aattgttatc
1681 cgctcacaat tccacacaac atacgagccg gaagcataaa gtgtaaagcc tggggtgcct
1741 aatgagtgag ctaactcaca ttaattgcgt tgcgctcact gcccgctttc cagtcgggaa
1801 acctgtcgtg ccagctgcat taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta
1861 ttgggcgctc ttccgcttcc tcgctcactg actcgctgcg ctcggtcgtt cggctgcggc
1921 gagcggtatc agctcactca aaggcggtaa tacggttatc cacagaatca ggggataacg
1981 caggaaagaa catgtgagca aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt
2041 tgctggcgtt tttccatagg ctccgccccc ctgacgagca tcacaaaaat cgacgctcaa
2101 gtcagaggtg gcgaaacccg acaggactat aaagatacca ggcgtttccc cctggaagct
2161 ccctcgtgcg ctctcctgtt ccgaccctgc cgcttaccgg atacctgtcc gcctttctcc
2221 cttcgggaag cgtggcgctt tctcatagct cacgctgtag gtatctcagt tcggtgtagg
2281 tcgttcgctc caagctgggc tgtgtgcacg aacccccccgt tcagcccgac cgctgcgcct
2341 tatccggtaa ctatcgtctt gagtccaacc cggtaagaca cgacttatcg ccactggcag
2401 cagccactgg taacaggatt agcagagcga ggtatgtagg cggtgctaca gagttcttga
2461 agtggtggcc taactacggc tacactagaa ggacagtatt tggtatctgc gctctgctga
2521 agccagttac cttcggaaaa agagttggta gctcttgatc cggcaaacaa accaccgctg
2581 gtagcggtgg ttttttttgtt tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag
2641 aagatccttt gatcttttct acggggtctg acgctcagtg gaacgaaaac tcacgttaag
2701 ggattttggt catgagatta tcaaaaagga tcttcaccta gatccttta aattaaaaat
2761 gaagttttaa atcaatctaa agtatatatg agtaaacttg gtctgacagt taccaatgct
2821 taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcatccata gttgcctgac
2881 tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa
2941 tgataccgcg agacccacgc tcaccggctc cagatttatc agcaataaac cagccagccg
3001 gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt
3061 gttgccggga agctagagta agtagttcgc cagttaatag tttgcgcaac gttgttgcca
3121 ttgctacagg catcgtggtg tcacgctcgt cgtttggtat ggcttcattc agctccggtt
3181 cccaacgatc aaggcgagtt acatgatccc ccatgttgtg caaaaaagcg gttagctcct
3241 tcggtcctcc gatcgttgtc agaagtaagt tggccgcagt gttatcactc atggttatgg
3301 cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct gtgactggtg
3361 agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg
3421 cgtcaatacg ggataatacc gcgccacata gcagaacttt aaaagtgctc atcattggaa
3481 aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt
3541 aacccactcg tgcacccaac tgatcttcag catcttttac tttcaccagc gtttctgggt
3601 gagcaaaaac aggaaggcaa aatgccgcaa aaaagggaat aagggcgaca cggaaatgtt
3661 gaatactcat actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca

3721 tgagcggata catatttgaa tgtatttaga aaaataaaca aataggggtt ccgcgcacat
3781 ttccccgaaa agtgccacct aaattgtaag cgttaatatt ttgttaaaat tcgcgttaaa
3841 tttttgttaa atcagctcat tttttaacca ataggccgaa atcggcaaaa tcccttataa
3901 atcaaaagaa tagaccgaga tagggttgag tgttgttcca gtttggaaca agagtccact
3961 attaaagaac gtggactcca acgtcaaagg gcgaaaaacc gtctatcagg gcgatggccc
4021 actacgtgaa ccatcaccct aatcaagttt tttggggtcg aggtgccgta aagcactaaa
4081 tcggaaccct aaagggagcc cccgatttag agcttgacgg ggaaagccgg cgaacgtggc
4141 gagaaaggaa gggaagaaag cgaaaggagc gggcgctagg gcgctggcaa gtgtagcggt
4201 cacgctgcgc gtaaccacca cacccgccgc gcttaatgcg ccgctacagg gcgcgtccca
4261 ttcgccattc aggctgcgca actgttggga agggcgatcg gtgcgggcct cttcgctatt
4321 acgccagctg gcgaaagggg gatgtgctgc aaggcgatta agttgggtaa cgccagggtt
4381 ttcccagtca cgacgttgta aaacgacggc cagtgagcgc gcgtaatacg actcactata
4441 gggcgaattg ggtaccgggc ccccctcga ggtcgacggt atcgataagc ttgatatcgt
4501 gtacaaaaaa gcaggcttta aaggaaccaa ttcagtcgac tggatccggt accaaggtcg
4561 ggcaggaaga gggcctattt cccatgattc cttcatattt gcatatacga tacaaggctg
4621 ttagagagat aattagaatt aatttgactg taaacacaaa gatattagta caaaatacgt
4681 gacgtagaaa gtaataattt cttgggtagt ttgcagtttt aaaattatgt tttaaaatgg
4741 actatcatat gcttaccgta acttgaaagt atttcgattt cttggcttta tatatcttgt
4801 ggaaaggacg aaacaccgtt taagagctat gctggaaaca gcatagcaag tttaaataag
4861 gctagtccgt tatcaacttg aaaaagtggc accgagtcgg tgcttcattt gcctgttccg
4921 tggccgacac

pBluescript II SK+ U6-lambda2-sgRNA(F+E) (SEQ ID NO: 34).

LOCUS Exported 3388 bp ds-DNA circular
DEFINITION synthetic circular DNA
SOURCE synthetic DNA construct
ORGANISM synthetic DNA construct
REFERENCE 1 (bases 1 to 3388)
FEATURES Location/Qualifiers
source 1..3388
/organism="synthetic DNA construct"
/mol_type="other DNA"
rep_origin 13..453
/direction=RIGHT
/label=F1 ori
CDS complement(460..528)
/label=LacZ alpha
primer_bind 599..616
/label=M13-fwd
primer_bind 626..645

/label=T7
promoter 757..1019
/label=U6 promoter
misc_feature 1020..1039
/label=lambda2_guideRNA
misc_feature 1041..1132
/label=sgRNA scaffold
primer_bind complement(1198..1217)
/label=T3
primer_bind complement(1235..1255)
/label=M13-rev
misc_binding complement(1261..1283)
/label=LacO
promoter complement(1288..1317)
/label=lac
rep_origin complement(1623..2251)
/direction=LEFT
/label=ColE1 origin
CDS complement(2403..3062)
/label=AmpR
promoter complement(3302..3330)
/label=Amp prom
ORIGIN
1 ctaaattgta agcgttaata ttttgttaaa attcgcgtta aatttttgtt aaatcagctc
61 atttttttaac caataggccg aaatcggcaa aatcccttat aaatcaaaag aatagaccga
121 gatagggttg agtgttgttc agtttggaa caagagtcca ctattaaaga acgtggactc
181 caacgtcaaa gggcgaaaaa ccgtctatca gggcgatggc ccactacgtg aaccatcacc
241 ctaatcaagt tttttggggt cgaggtgccg taaagcacta aatcggaacc ctaaagggag
301 cccccgattt agagcttgac ggggaaagcc ggcgaacgtg gcgagaaagg aagggaagaa
361 agcgaaagga gcgggcgcta gggcgctggc aagtgtagcg gtcacgctgc gcgtaaccac
421 cacacccgcc gcgcttaatg cgccgctaca gggcgcgtcc cattcgccat tcaggctgcg
481 caactgttgg gaagggcgat cggtgcgggc ctcttcgcta ttacgccagc tggcgaaagg
541 gggatgtgct gcaaggcgat taagttgggt aacgccaggg ttttcccagt cacgacgttg
601 taaaacgacg gccagtgagc gcgcgtaata cgactcacta tagggcgaat tgggtaccgg
661 gccccccctc gaggtcgacg gtatcgataa gcttgatatc gtgtacaaaa aagcaggctt
721 taaaggaacc aattcagtcg actggatccg gtaccaaggt cgggcaggaa gagggcctat
781 ttcccatgat tccttcatat ttgcatatac gatacaaggc tgttagagag ataattagaa
841 ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga aagtaataat
901 ttcttgggta gtttgcagtt ttaaaattat gttttaaaat ggactatcat atgcttaccg
961 taacttgaaa gtatttcgat ttcttggctt tatatatctt gtggaaagga cgaaacaccg
1021 tgataagtgg aatgccatgg tttaagagct atgctggaaa cagcatagca agtttaaata

1081 aggctagtcc gttatcaact tgaaaaagtg gcaccgagtc ggtgcttttt ttcctgcagc
1141 ccgggggatc cactagttct agagcggccg ccaccgcggt ggagctccag cttttgttcc
1201 ctttagtgag ggttaattgc gcgcttggcg taatcatggt catagctgtt tcctgtgtga
1261 aattgttatc cgctcacaat tccacacaac atacgagccg gaagcataaa gtgtaaagcc
1321 tggggtgcct aatgagtgag ctaactcaca ttaattgcgt tgcgctcact gcccgctttc
1381 cagtcgggaa acctgtcgtg ccagctgcat taatgaatcg gccaacgcgc ggggagaggc
1441 ggtttgcgta ttgggcgctc ttccgcttcc tcgctcactg actcgctgcg ctcggtcgtt
1501 cggctgcggc gagcggtatc agctcactca aaggcggtaa tacggttatc cacagaatca
1561 ggggataacg caggaaagaa catgtgagca aaaggccagc aaaaggccag gaaccgtaaa
1621 aaggccgcgt tgctggcgtt tttccatagg ctccgccccc ctgacgagca tcacaaaaat
1681 cgacgctcaa gtcagaggtg gcgaaacccg acaggactat aaagatacca ggcgtttccc
1741 cctggaagct ccctcgtgcg ctctcctgtt ccgaccctgc cgcttaccgg atacctgtcc
1801 gcctttctcc cttcgggaag cgtggcgctt tctcatagct cacgctgtag gtatctcagt
1861 tcggtgtagg tcgttcgctc caagctgggc tgtgtgcacg aaccccccgt tcagcccgac
1921 cgctgcgcct tatccggtaa ctatcgtctt gagtccaacc cggtaagaca cgacttatcg
1981 ccactggcag cagccactgg taacaggatt agcagagcga ggtatgtagg cggtgctaca
2041 gagttcttga agtggtggcc taactacggc tacactagaa ggacagtatt tggtatctgc
2101 gctctgctga agccagttac cttcggaaaa agagttggta gctcttgatc cggcaaacaa
2161 accaccgctg gtagcggtgg ttttttgtt tgcaagcagc agattacgcg cagaaaaaaa
2221 ggatctcaag aagatccttt gatcttttct acggggtctg acgctcagtg gaacgaaaac
2281 tcacgttaag ggattttggt catgagatta tcaaaaagga tcttcaccta gatccttta
2341 aattaaaaat gaagttttaa atcaatctaa agtatatatg agtaaacttg gtctgacagt
2401 taccaatgct taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcatccata
2461 gttgcctgac tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc
2521 agtgctgcaa tgataccgcg agacccacgc tcaccggctc cagatttatc agcaataaac
2581 cagccagccg gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag
2641 tctattaatt gttgccggga agctagagta agtagttcgc cagttaatag tttgcgcaac
2701 gttgttgcca ttgctacagg catcgtggtg tcacgctcgt cgtttggtat ggcttcattc
2761 agctccggtt cccaacgatc aaggcgagtt acatgatccc ccatgttgtg caaaaaagcg
2821 gttagctcct tcggtcctcc gatcgttgtc agaagtaagt tggccgcagt gttatcactc
2881 atggttatgg cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct
2941 gtgactggtg agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc
3001 tcttgcccgg cgtcaatacg ggataatacc gcgccacata gcagaacttt aaaagtgctc
3061 atcattggaa aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc
3121 agttcgatgt aacccactcg tgcacccaac tgatcttcag catcttttac tttcaccagc
3181 gtttctgggt gagcaaaaac aggaaggcaa aatgccgcaa aaaagggaat aagggcgaca
3241 cggaaatgtt gaatactcat actcttcctt tttcaatatt attgaagcat ttatcagggt
3301 tattgtctca tgagcggata catatttgaa tgtatttaga aaaataaaca aatagggggtt
3361 ccgcgcacat ttccccgaaa agtgccac

EGFP_spacerless_SaCas9_sgRNA (SEQ ID NO: 47)

LOCUS Exported 4921 bp ds-DNA circular
DEFINITION synthetic circular DNA
SOURCE synthetic DNA construct
ORGANISM recombinant plasmid
REFERENCE 1 (bases 1 to 4921)
FEATURES Location/Qualifiers
source 1..4921
/organism="recombinant plasmid"
/mol_type="other DNA"
primer_bind 1..40
/label=EF1a_Gibson_F
primer_bind 1..20
/label=Primer 2
misc_feature 1..7
/label=sgRNA scaffold_termination
promoter 21..566
/label=EF1a promoter
primer_bind complement(554..591)
/label=EF1a_Gibson_R
CDS 572..1282
/codon_start=1
/product="monomeric derivative of DsRed fluorescent protein
(Shaner et al., 2004)"
/label=mCherry
/note="mammalian codon-optimized"

/translation="MVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEG
TQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNF
EDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALK
GEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERA
EGRHSTGGMDELYK"
primer_bind 572..591
/label=mCherry_BGH_F
primer_bind complement(1259..1306)
/label=Primer 1
primer_bind complement(1259..1286)
/label=mCherry_P2A_Gib_R
primer_bind complement(1259..1282)
/label=mCherry_HindIII_R
misc_feature 1283..1306

/label=Gibson Overlap
primer_bind 1283..1301
/label=mCherry_P2A_Gib_F
polyA_signal 1330..1554
/label=bGH poly(A) signal
/note="bovine growth hormone polyadenylation signal"
primer_bind complement(1535..1573)
/label=mCherry_BGH_Gib_R
primer_bind complement(1535..1554)
/label=mCherry_BGH_R
primer_bind complement(1536..1555)
/label=bGH_NotI_R
primer_bind complement(1558..1573)
/label=SK primer
/note="common sequencing primer, one of multiple similar
variants"
primer_bind complement(1608..1627)
/label=T3
primer_bind complement(1645..1665)
/label=M13-rev
misc_binding complement(1671..1693)
/label=LacO
promoter complement(1698..1727)
/label=lac
rep_origin complement(2033..2661)
/direction=LEFT
/label=ColE1 origin
CDS complement(2813..3472)
/label=AmpR
promoter complement(3712..3740)
/label=Amp prom
rep_origin 3811..4251
/direction=RIGHT
/label=F1 ori
CDS complement(4258..4326)
/label=LacZ alpha
primer_bind 4397..4414
/label=M13-fwd
primer_bind 4424..4443
/label=T7
promoter 4555..4817

/label=U6 promoter
primer_bind 4798..4843
/label=NS_EGFP_SaCas9_F
primer_bind complement(4803..4817)
/label=Primer 3
primer_bind complement(4803..4817)
/label=extension_gibson_R
primer_bind 4804..4843
/label=50bp_EGFP_SaCas9_F
misc_RNA 4819..4894
/label=Sa gRNA scaffold
/note="guide RNA scaffold for the Staphylococcus aureus CRISPR/Cas9 system"
primer_bind complement(join(4877..4921,1..20))
/label=EGFP_SaCas9_RC_ex_R
primer_bind complement(join(4877..4921,1..20))
/label=EGFP_SaCas9_ex_R
misc_feature 4895..4900
/label=Linker
misc_feature 4901..4921
/label=EGFP extension
primer_bind 4901..4921
/label=RNA target with T7 Promoter Sequence (for IVT)
ORIGIN
1 tttttttcct gcagcccggg aaggatctgc gatcgctccg gtgcccgtca gtgggcagag
61 cgcacatcgc ccacagtccc cgagaagttg gggggagggg tcggcaattg aacgggtgcc
121 tagagaaggt ggcgcggggt aaactgggaa agtgatgtcg tgtactggct ccgccttttt
181 cccgagggtg ggggagaacc gtatataagt gcagtagtcg ccgtgaacgt tcttttcgc
241 aacgggtttg ccgccagaac acagctgaag cttcgagggg ctcgcatctc tccttcacgc
301 gcccgccgcc ctacctgagg ccgccatcca cgccggttga gtcgcgttct gccgcctccc
361 gcctgtggtg cctcctgaac tgcgtccgcc gtctaggtaa gtttaaagct caggtcgaga
421 ccgggccttt gtccggcgct cccttggagc ctacctagac tcagccggct ctccacgctt
481 tgcctgaccc tgcttgctca actctacgtc tttgtttcgt tttctgttct gcgccgttac
541 agatccaagc tgtgaccggc gcctacgcta gatggtgagc aagggcgagg aggataacat
601 ggccatcatc aaggagttca tgcgcttcaa ggtgcacatg gagggctccg tgaacggcca
661 cgagttcgag atcgagggcg agggcgaggg ccgcccctac gagggcaccc agaccgccaa
721 gctgaaggtg accaagggtg gcccccctgcc cttcgcctgg gacatcctgt cccctcagtt
781 catgtacggc tccaaggcct acgtgaagca ccccgccgac atccccgact acttgaagct
841 gtccttcccc gagggcttca agtgggagcg cgtgatgaac ttcgaggacg gcggcgtggt
901 gaccgtgacc caggactcct ccctgcagga cggcgagttc atctacaagg tgaagctgcg
961 cggcaccaac ttcccctccg acggccccgt aatgcagaag aagaccatgg gctgggaggc

1021 ctcctccgag cggatgtacc ccgaggacgg cgccctgaag ggcgagatca agcagaggct
1081 gaagctgaag gacggcggcc actacgacgc tgaggtcaag accacctaca aggccaagaa
1141 gcccgtgcag ctgcccggcg cctacaacgt caacatcaag ttggacatca cctcccacaa
1201 cgaggactac accatcgtgg aacagtacga acgcgccgag ggccgccact ccaccggcgg
1261 catggacgag ctgtacaagt aatccgagct cggtaccaag cttaagttta aaccgctgat
1321 cagcctcgac tgtgccttct agttgccagc catctgttgt ttgcccctcc cccgtgcctt
1381 ccttgaccct ggaaggtgcc actcccactg tcctttccta ataaaatgag gaaattgcat
1441 cgcattgtct gagtaggtgt cattctattc tgggggggtgg ggtggggcag gacagcaagg
1501 gggaggattg ggaagacaat agcaggcatg ctggggatgc ggtgggctct atgggggatc
1561 cactagttct agagcggccg ccaccgcggt ggagctccag cttttgttcc ctttagtgag
1621 ggttaattgc gcgcttggcg taatcatggt catagctgtt tcctgtgtga aattgttatc
1681 cgctcacaat tccacacaac atacgagccg gaagcataaa gtgtaaagcc tggggtgcct
1741 aatgagtgag ctaactcaca ttaattgcgt tgcgctcact gcccgctttc cagtcgggaa
1801 acctgtcgtg ccagctgcat taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta
1861 ttgggcgctc ttccgcttcc tcgctcactg actcgctgcg ctcggtcgtt cggctgcggc
1921 gagcggtatc agctcactca aaggcggtaa tacggttatc cacagaatca ggggataacg
1981 caggaaagaa catgtgagca aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt
2041 tgctggcgtt tttccatagg ctccgccccc ctgacgagca tcacaaaaat cgacgctcaa
2101 gtcagaggtg gcgaaacccg acaggactat aaagatacca ggcgtttccc cctggaagct
2161 ccctcgtgcg ctctcctgtt ccgaccctgc cgcttaccgg atacctgtcc gcctttctcc
2221 cttcgggaag cgtggcgctt tctcatagct cacgctgtag gtatctcagt tcggtgtagg
2281 tcgttcgctc caagctgggc tgtgtgcacg aacccccccgt tcagcccgac cgctgcgcct
2341 tatccggtaa ctatcgtctt gagtccaacc cggtaagaca cgacttatcg ccactggcag
2401 cagccactgg taacaggatt agcagagcga ggtatgtagg cggtgctaca gagttcttga
2461 agtggtggcc taactacggc tacactagaa ggacagtatt tggtatctgc gctctgctga
2521 agccagttac cttcggaaaa agagttggta gctcttgatc cggcaaacaa accaccgctg
2581 gtagcggtgg ttttttttgtt tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag
2641 aagatccttt gatcttttct acggggtctg acgctcagtg gaacgaaaac tcacgttaag
2701 ggattttggt catgagatta tcaaaaagga tcttcaccta gatcctttta aattaaaaat
2761 gaagttttaa atcaatctaa agtatatatg agtaaacttg gtctgacagt taccaatgct
2821 taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcatccata gttgcctgac
2881 tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa
2941 tgataccgcg agacccacgc tcaccggctc cagatttatc agcaataaac cagccagccg
3001 gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt
3061 gttgccggga agctagagta agtagttcgc cagttaatag tttgcgcaac gttgttgcca
3121 ttgctacagg catcgtggtg tcacgctcgt cgtttggtat ggcttcattc agctccggtt
3181 cccaacgatc aaggcgagtt acatgatccc ccatgttgtg caaaaaagcg gttagctcct
3241 tcggtcctcc gatcgttgtc agaagtaagt tggccgcagt gttatcactc atggttatgg
3301 cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct gtgactggtg
3361 agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg
3421 cgtcaatacg ggataatacc gcgccacata gcagaacttt aaaagtgctc atcattggaa

3481 aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt
3541 aacccactcg tgcacccaac tgatcttcag catcttttac tttcaccagc gtttctgggt
3601 gagcaaaaac aggaaggcaa aatgccgcaa aaaagggaat aagggcgaca cggaaatgtt
3661 gaatactcat actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca
3721 tgagcggata catatttgaa tgtatttaga aaaataaaca aataggggtt ccgcgcacat
3781 ttccccgaaa agtgccacct aaattgtaag cgttaatatt ttgttaaaat tcgcgttaaa
3841 tttttgttaa atcagctcat tttttaacca ataggccgaa atcggcaaaa tcccttataa
3901 atcaaaagaa tagaccgaga tagggttgag tgttgttcca gtttggaaca agagtccact
3961 attaaagaac gtggactcca acgtcaaagg gcgaaaaacc gtctatcagg gcgatggccc
4021 actacgtgaa ccatcaccct aatcaagttt tttggggtcg aggtgccgta aagcactaaa
4081 tcggaaccct aaagggagcc cccgatttag agcttgacgg ggaaagccgg cgaacgtggc
4141 gagaaaggaa gggaagaaag cgaaaggagc gggcgctagg gcgctggcaa gtgtagcggt
4201 cacgctgcgc gtaaccacca cacccgccgc gcttaatgcg ccgctacagg gcgcgtccca
4261 ttcgccattc aggctgcgca actgttggga agggcgatcg gtgcgggcct cttcgctatt
4321 acgccagctg gcgaaagggg gatgtgctgc aaggcgatta agttgggtaa cgccagggtt
4381 ttcccagtca cgacgttgta aaacgacggc cagtgagcgc gcgtaatacg actcactata
4441 gggcgaattg ggtaccgggc ccccctcga ggtcgacggt atcgataagc ttgatatcgt
4501 gtacaaaaaa gcaggcttta aaggaaccaa ttcagtcgac tggatccggt accaaggtcg
4561 ggcaggaaga gggcctattt cccatgattc cttcatattt gcatatacga tacaaggctg
4621 ttagagagat aattagaatt aatttgactg taaacacaaa gatattagta caaaatacgt
4681 gacgtagaaa gtaataattt cttgggtagt ttgcagtttt aaaattatgt tttaaaatgg
4741 actatcatat gcttaccgta acttgaaagt atttcgattt cttggcttta tatatcttgt
4801 ggaaaggacg aaacaccggt tatagtactc tggaaacaga atctactata acaaggcaaa
4861 atgccgtgtt tatctcgtca acttgttggc gagattcatt gtgtcggcca cggaacaggc
4921 a

ADAR2_E488Q_dSaCas9_pCDNA3_1 (SEQ ID NO: 48)

LOCUS Exported 9842 bp ds-DNA circular
DEFINITION synthetic circular DNA
SOURCE synthetic DNA construct
ORGANISM recombinant plasmid
REFERENCE 1 (bases 1 to 9842)
FEATURES Location/Qualifiers
source 1..9842
/organism="recombinant plasmid"
/mol_type="other DNA"
primer_bind complement(213..234)
/label=pCDNA3_CMV_out_R
enhancer 235..614
/label=CMV enhancer

/note="human cytomegalovirus immediate early enhancer"
promoter 615..818
/label=CMV promoter
/note="human cytomegalovirus (CMV) immediate early promoter"
promoter 863..881
/label=T7 promoter
/note="promoter for bacteriophage T7 RNA polymerase"
primer_bind 927..985
/label=H1-ADAR-XTEN_F
misc_feature 927..954
/label=Homology 1_pCDNA3.1
CDS 961..2100
/codon_start=1
/label=ADARB1(E488Q)_Catalytic Domain

/translation="MLADAVSRLVLGKFGDLTDNFSSPHARRKVLAGVVMTTGTDVKDA
KVISVSTGTKCINGEYMSDRGLALNDCHAEIISRRSLLRFLYTQLELYLNNKDDQKRSI
FQKSERGGFRLKENVQFHLYISTSPCGDARIFSPHEPILEEPADRHPNRKARGQLRTKI
ESGQGTIPVRSNASIQTWDGVLQGERLLTMSCSDKIARWNVVGIQGSLLSIFVEPIYFS
SIILGSLYHGDHLSRAMYQRISNIEDLPPLYTLNKPLLSGISNAEARQPGKAPNFSVNW
TVGDSAIEVINATTGKDELGRASRLCKHALYCRWMRVHGKVPSHLLRSKITKPNVYHES
KLAAKEYQAAKARLFTAFIKAGLGAWVEKPTEQDQFSLTP"
primer_bind 961..982
/label=Primer 4
primer_bind 1111..1138
/label=Primer 1
primer_bind 1440..1478
/label=E488Q_Mutagenesis_F
primer_bind complement(1440..1478)
/label=E488Q_Mutagenesis_R
primer_bind complement(2080..2112)
/label=ADAR2DD_GS_R
primer_bind complement(2080..2100)
/label=Primer 5
primer_bind 2086..2132
/label=SaCas9_Gib_F
misc_feature 2101..2112
/label=GS_linker
misc_feature 2113..5268
/label=dSaCas9(D10A,N580A)

primer_bind complement(5245..5268)
/label=SaCas9_Gib_R
primer_bind 5249..5289
/label=SaCas9_HA_F
primer_bind 5269..5290
/label=ADAR2_CD_Inverse_F
CDS 5272..5298
/codon_start=1
/product="HA (human influenza hemagglutinin) epitope tag"
/label=HA
/translation="YPYDVPDYA"
primer_bind complement(5290..5312)
/label=AXC_NLSout_NESin_R
primer_bind complement(5290..5310)
/label=NLS_out_R
CDS 5317..5337
/codon_start=1
/product="nuclear localization signal of SV40 large T
antigen"
/label=SV40 NLS
/translation="PKKKRKV"
CDS 5344..5364
/codon_start=1
/product="nuclear localization signal of SV40 large T
antigen"
/label=SV40 NLS
/translation="PKKKRKV"
primer_bind complement(5349..5408)
/label=XTEN-Cas9-H2_R
primer_bind complement(5349..5393)
/label=Primer 7
primer_bind 5363..5387
/label=NLS_out_NES_full_F
primer_bind 5365..5387
/label=AXC_NLSout_NESin_F
misc_feature 5374..5408
/label=Homology 2_pCDNA3.1
primer_bind 5374..5392
/label=pCDNA3_CMV_out_F
primer_bind 5395..5418
/label=bGH_HindIII_F

polyA_signal 5442..5666
/label=bGH poly(A) signal
/note="bovine growth hormone polyadenylation signal"
primer_bind complement(5648..5666)
/label=bGH_NotI_R
rep_origin 5712..6140
/direction=RIGHT
/label=f1 ori
/note="f1 bacteriophage origin of replication; arrow
indicates direction of (+) strand synthesis"
promoter 6154..6483
/label=SV40 promoter
/note="SV40 enhancer and early promoter"
rep_origin 6334..6469
/label=SV40 ori
/note="SV40 origin of replication"
CDS 6550..7344
/codon_start=1
/gene="aph(3')-II (or nptII)"
/product="aminoglycoside phosphotransferase from Tn5"
/label=NeoR/KanR
/note="confers resistance to neomycin, kanamycin, and G418
(Geneticin(R))"

/translation="MIEQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQGRP
VLFVKTDLSGALNELQDEAARLSWLATTGVPCAAVLDVVTEAGRDWLLLGEVPGQDLLS
SHLAPAEKVSIMADAMRRLHTLDPATCPFDHQAKHRIERARTRMEAGLVDQDDLDEEHQ
GLAPAELFARLKARMPDGEDLVVTHGDACLPNIMVENGRFSGFIDCGRLGVADRYQDIA
LATRDIAEELGGEWADRFLVLYGIAAPDSQRIAFYRLLDEFF"
polyA_signal 7518..7639
/label=SV40 poly(A) signal
/note="SV40 polyadenylation signal"
primer_bind complement(7688..7704)
/label=M13 rev
/note="common sequencing primer, one of multiple similar
variants"
protein_bind 7712..7728
/label=lac operator
/bound_moiety="lac repressor encoded by lacI"
/note="The lac repressor binds to the lac operator to
inhibit transcription in E. coli. This inhibition can be

relieved by adding lactose or
isopropyl-beta-D-thiogalactopyranoside (IPTG)."
promoter complement(7736..7766)
/label=lac promoter
/note="promoter for the E. coli lac operon"
protein_bind 7781..7802
/label=CAP binding site
/bound_moiety="E. coli catabolite activator protein"
/note="CAP binding activates transcription in the presence
of cAMP."
rep_origin complement(8090..8675)
/direction=LEFT
/label=ori
/note="high-copy-number ColE1/pMB1/pBR322/pUC origin of
replication"
CDS complement(8846..9706)
/codon_start=1
/gene="bla"
/product="beta-lactamase"
/label=AmpR
/note="confers resistance to ampicillin, carbenicillin, and
related antibiotics"

/translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGYI
ELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVEYS
PVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRLDRW
EPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSA
LPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGAS
LIKHW"
promoter complement(9707..9811)
/gene="bla"
/label=AmpR promoter
ORIGIN
1 gacggatcgg gagatctccc gatcccctat ggtgcactct cagtacaatc tgctctgatg
61 ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgct gagtagtgcg
121 cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc
181 ttagggttag gcgttttgcg ctgcttcgcg atgtacgggc cagatatacg cgttgacatt
241 gattattgac tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata
301 tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc
361 cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc
421 attgacgtca atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt

481 atcatatgcc aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt
541 atgcccagta catgacctta tgggactttc ctacttggca gtacatctac gtattagtca
601 tcgctattac catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg
661 actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc
721 aaaatcaacg ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg
781 gtaggcgtgt acggtgggag gtctatataa gcagagctct ctggctaact agagaaccca
841 ctgcttactg gcttatcgaa attaatacga ctcactatag ggagacccaa gctggctagc
901 gtttaaacgg gccctctaga ctcgagcggc cgccactgtg ctggatatct gcagagaacc
961 atgttagctg acgctgtctc acgcctggtc ctgggtaagt ttggtgacct gaccgacaac
1021 ttctcctccc ctcacgctcg cagaaaagtg ctggctggag tcgtcatgac aacaggcaca
1081 gatgttaaag atgccaaggt gataagtgtt tctacaggaa caaaatgtat taatggtgaa
1141 tacatgagtg atcgtggcct tgcattaaat gactgccatg cagaaataat atctcggaga
1201 tccttgctca gatttcttta tacacaactt gagctttact taaataacaa agatgatcaa
1261 aaaagatcca tctttcagaa atcagagcga ggggggttta ggctgaagga gaatgtccag
1321 tttcatctgt acatcagcac ctctccctgt ggagatgcca gaatcttctc accacatgag
1381 ccaatcctgg aagaaccagc agatagacac ccaaatcgta aagcaagagg acagctacgg
1441 accaaaatag agtctggtca ggggacgatt ccagtgcgct ccaatgcgag catccaaacg
1501 tgggacgggg tgctgcaagg ggagcggctg ctcaccatgt cctgcagtga caagattgca
1561 cgctggaacg tggtgggcat ccagggatcc ctgctcagca ttttcgtgga gcccatttac
1621 ttctcgagca tcatcctggg cagcctttac cacggggacc acctttccag ggccatgtac
1681 cagcggatct ccaacataga ggacctgcca cctctctaca ccctcaacaa gcctttgctc
1741 agtggcatca gcaatgcaga agcacggcag ccagggaagg cccccaactt cagtgtcaac
1801 tggacggtag gcgactccgc tattgaggtc atcaacgcca cgactgggaa ggatgagctg
1861 ggccgcgcgt cccgcctgtg taagcacgcg ttgtactgtc gctggatgcg tgtgcacggc
1921 aaggttccct cccacttact acgctccaag attaccaagc ccaacgtgta ccatgagtcc
1981 aagctggcgg caaaggagta ccaggccgcc aaggcgcgtc tgttcacagc cttcatcaag
2041 gcggggctgg gggcctgggt ggagaagccc accgagcagg accagttctc actcacgccc
2101 ggatccggat ccaagcggaa ctacatcctg ggcctggcca tcggcatcac cagcgtgggc
2161 tacggcatca tcgactacga gacacgggac gtgatcgatg ccggcgtgcg gctgttcaaa
2221 gaggccaacg tggaaaacaa cgagggcagg cggagcaaga gaggcgccag aaggctgaag
2281 cggcggaggc ggcatagaat ccagagagtg aagaagctgc tgttcgacta caacctgctg
2341 accgaccaca gcgagctgag cggcatcaac ccctacgagg ccagagtgaa gggcctgagc
2401 cagaagctga gcgaggaaga gttctctgcc gccctgctgc acctggccaa gagaagaggc
2461 gtgcacaacg tgaacgaggt ggaagaggac accggcaacg agctgtccac caaagagcag
2521 atcagccgga acagcaaggc cctggaagag aaatacgtgg ccgaactgca gctggaacgg
2581 ctgaagaaag acggcgaagt gcggggcagc atcaacagat tcaagaccag cgactacgtg
2641 aaagaagcca aacagctgct gaaggtgcag aaggcctacc accagctgga ccagagcttc
2701 atcgacacct acatcgacct gctggaaacc cggcggacct actatgaggg acctggcgag
2761 ggcagcccct tcggctggaa ggacatcaaa gaatggtacg agatgctgat gggccactgc
2821 acctacttcc ccgaggaact gcggagcgtg aagtacgcct acaacgccga cctgtacaac
2881 gccctgaacg acctgaacaa tctcgtgatc accagggacg agaacgagaa gctggaatat

2941 tacgagaagt tccagatcat cgagaacgtg ttcaagcaga agaagaagcc caccctgaag
3001 cagatcgcca aagaaatcct cgtgaacgaa gaggatatta agggctacag agtgaccagc
3061 accggcaagc ccgagttcac caacctgaag gtgtaccacg acatcaagga cattaccgcc
3121 cggaaagaga ttattgagaa cgccgagctg ctggatcaga ttgccaagat cctgaccatc
3181 taccagagca gcgaggacat ccaggaagaa ctgaccaatc tgaactccga gctgacccag
3241 gaagagatcg agcagatctc taatctgaag ggctataccg gcacccacaa cctgagcctg
3301 aaggccatca acctgatcct ggacgagctg tggcacacca acgacaacca gatcgctatc
3361 ttcaaccggc tgaagctggt gcccaagaag gtggacctgt cccagcagaa agagatcccc
3421 accaccctgg tggacgactt catcctgagc cccgtcgtga agagaagctt catccagagc
3481 atcaaagtga tcaacgccat catcaagaag tacggcctgc ccaacgacat cattatcgag
3541 ctggcccgcg agaagaactc caaggacgcc cagaaaatga tcaacgagat gcagaagcgg
3601 aaccggcaga ccaacgagcg gatcgaggaa atcatccgga ccaccggcaa agagaacgcc
3661 aagtacctga tcgagaagat caagctgcac gacatgcagg aaggcaagtg cctgtacagc
3721 ctggaagcca tccctctgga agatctgctg aacaacccct tcaactatga ggtggaccac
3781 atcatcccca gaagcgtgtc cttcgacaac agcttcaaca acaaggtgct cgtgaagcag
3841 gaagaagcca gcaagaaggg caaccggacc ccattccagt acctgagcag cagcgacagc
3901 aagatcagct acgaaacctt caagaagcac atcctgaatc tggccaaggg caagggcaga
3961 atcagcaaga ccaagaaaga gtatctgctg gaagaacggg acatcaacag gttctccgtg
4021 cagaaagact tcatcaaccg gaacctggtg gataccagat acgccaccag aggcctgatg
4081 aacctgctgc ggagctactt cagagtgaac aacctggacg tgaaagtgaa gtccatcaat
4141 ggcggcttca ccagctttct gcggcggaag tggaagttta agaaagagcg gaacaagggg
4201 tacaagcacc acgccgagga cgccctgatc attgccaacg ccgatttcat cttcaaagag
4261 tggaagaaac tggacaaggc caaaaaagtg atggaaaacc agatgttcga ggaaaagcag
4321 gccgagagca tgcccgagat cgaaaccgag caggagtaca aagagatctt catcaccccc
4381 caccagatca agcacattaa ggacttcaag gactacaagt acagccaccg ggtggacaag
4441 aagcctaata gagagctgat taacgacacc ctgtactcca cccggaagga cgacaagggc
4501 aacaccctga tcgtgaacaa tctgaacggc ctgtacgaca aggacaatga caagctgaaa
4561 aagctgatca acaagagccc cgaaaagctg ctgatgtacc accacgaccc ccagacctac
4621 cagaaactga agctgattat ggaacagtac ggcgacgaga agaatcccct gtacaagtac
4681 tacgaggaaa ccgggaacta cctgaccaag tactccaaaa aggacaacgg ccccgtgatc
4741 aagaagatta agtattacgg caacaaactg aacgcccatc tggacatcac cgacgactac
4801 cccaacagca gaaacaaggt cgtgaagctg tccctgaagc cctacagatt cgacgtgtac
4861 ctggacaatg gcgtgtacaa gttcgtgacc gtgaagaatc tggatgtgat caaaaaagaa
4921 aactactacg aagtgaatag caagtgctat gaggaagcta agaagctgaa gaagatcagc
4981 aaccaggccg agtttatcgc ctccttctac aacaacgatc tgatcaagat caacggcgag
5041 ctgtatagag tgatcggcgt gaacaacgac ctgctgaacc ggatcgaagt gaacatgatc
5101 gacatcacct accgcgagta cctggaaaac atgaacgaca agaggccccc caggatcatt
5161 aagacaatcg cctccaagac ccagagcatt aagaagtaca gcacagacat tctgggcaac
5221 ctgtatgaag tgaaatctaa gaagcaccct cagatcatca aaaagggcgc ctatccctat
5281 gacgtgcccg attatgccag cctgggcagc ggctccccca agaaaaaacg caaggtggaa
5341 gatcctaaga aaaagcggaa agtggacgtg taaccaccac actggactag tggatccgag

5401 ctcggtacca agcttaagtt taaaccgctg atcagcctcg actgtgcctt ctagttgcca
5461 gccatctgtt gtttgcccct cccccgtgcc ttccttgacc ctggaaggtg ccactcccac
5521 tgtcctttcc taataaaatg aggaaattgc atcgcattgt ctgagtaggt gtcattctat
5581 tctgggggggt ggggtggggc aggacagcaa gggggaggat tgggaagaca atagcaggca
5641 tgctggggat gcggtgggct ctatggcttc tgaggcggaa agaaccagct ggggctctag
5701 ggggtatccc cacgcgccct gtagcggcgc attaagcgcg gcgggtgtgg tggttacgcg
5761 cagcgtgacc gctacacttg ccagcgccct agcgcccgct cctttcgctt tcttcccttc
5821 ctttctcgcc acgttcgccg gctttccccg tcaagctcta aatcgggggc tccctttagg
5881 gttccgattt agtgctttac ggcacctcga ccccaaaaaa cttgattagg gtgatggttc
5941 acgtagtggg ccatcgccct gatagacggt ttttcgccct ttgacgttgg agtccacgtt
6001 ctttaatagt ggactcttgt tccaaactgg aacaacactc aaccctatct cggtctattc
6061 ttttgattta taagggattt tgccgatttc ggcctattgg ttaaaaaatg agctgattta
6121 acaaaaattt aacgcgaatt aattctgtgg aatgtgtgtc agttagggtg tggaaagtcc
6181 ccaggctccc cagcaggcag aagtatgcaa agcatgcatc tcaattagtc agcaaccagg
6241 tgtggaaagt ccccaggctc cccagcaggc agaagtatgc aaagcatgca tctcaattag
6301 tcagcaacca tagtcccgcc cctaactccg cccatcccgc ccctaactcc gcccagttcc
6361 gcccattctc cgccccatgg ctgactaatt ttttttattt atgcagaggc cgaggccgcc
6421 tctgcctctg agctattcca gaagtagtga ggaggctttt ttggaggcct aggcttttgc
6481 aaaaagctcc cgggagcttg tatatccatt ttcggatctg atcaagagac aggatgagga
6541 tcgtttcgca tgattgaaca agatggattg cacgcaggtt ctccggccgc ttgggtggag
6601 aggctattcg gctatgactg ggcacaacag acaatcggct gctctgatgc cgccgtgttc
6661 cggctgtcag cgcaggggcg cccggttctt tttgtcaaga ccgacctgtc cggtgccctg
6721 aatgaactgc aggacgaggc agcgcggcta tcgtggctgg ccacgacggg cgttccttgc
6781 gcagctgtgc tcgacgttgt cactgaagcg ggaagggact ggctgctatt gggcgaagtg
6841 ccggggcagg atctcctgtc atctcacctt gctcctgccg agaaagtatc catcatggct
6901 gatgcaatgc ggcggctgca tacgcttgat ccggctacct gcccattcga ccaccaagcg
6961 aaacatcgca tcgagcgagc acgtactcgg atggaagccg gtcttgtcga tcaggatgat
7021 ctggacgaag agcatcaggg gctcgcgcca gccgaactgt tcgccaggct caaggcgcgc
7081 atgcccgacg gcgaggatct cgtcgtgacc catggcgatg cctgcttgcc gaatatcatg
7141 gtggaaaatg gccgcttttc tggattcatc gactgtggcc ggctgggtgt ggcggaccgc
7201 tatcaggaca tagcgttggc tacccgtgat attgctgaag agcttggcgg cgaatgggct
7261 gaccgcttcc tcgtgcttta cggtatcgcc gctcccgatt cgcagcgcat cgccttctat
7321 cgccttcttg acgagttctt ctgagcggga ctctggggtt cgaaatgacc gaccaagcga
7381 cgcccaacct gccatcacga gatttcgatt ccaccgccgc cttctatgaa aggttgggct
7441 tcggaatcgt tttccgggac gccggctgga tgatcctcca gcgcggggat ctcatgctgg
7501 agttcttcgc ccaccccaac ttgtttattg cagcttataa tggttacaaa taaagcaata
7561 gcatcacaaa tttcacaaat aaagcatttt tttcactgca ttctagttgt ggtttgtcca
7621 aactcatcaa tgtatcttat catgtctgta taccgtcgac ctctagctag agcttggcgt
7681 aatcatggtc atagctgttt cctgtgtgaa attgttatcc gctcacaatt ccacacaaca
7741 tacgagccgg aagcataaag tgtaaagcct ggggtgccta atgagtgagc taactcacat
7801 taattgcgtt gcgctcactg cccgctttcc agtcgggaaa cctgtcgtgc cagctgcatt

7861 aatgaatcgg ccaacgcgcg gggagaggcg gtttgcgtat tgggcgctct tccgcttcct
7921 cgctcactga ctcgctgcgc tcggtcgttc ggctgcggcg agcggtatca gctcactcaa
7981 aggcggtaat acggttatcc acagaatcag gggataacgc aggaaagaac atgtgagcaa
8041 aaggccagca aaaggccagg aaccgtaaaa aggccgcgtt gctggcgttt ttccataggc
8101 tccgcccccc tgacgagcat cacaaaaatc gacgctcaag tcagaggtgg cgaaacccga
8161 caggactata aagataccag gcgtttcccc ctggaagctc cctcgtgcgc tctcctgttc
8221 cgaccctgcc gcttaccgga tacctgtccg cctttctccc ttcgggaagc gtggcgcttt
8281 ctcatagctc acgctgtagg tatctcagtt cggtgtaggt cgttcgctcc aagctgggct
8341 gtgtgcacga acccccccgtt cagcccgacc gctgcgcctt atccggtaac tatcgtcttg
8401 agtccaaccc ggtaagacac gacttatcgc cactggcagc agccactggt aacaggatta
8461 gcagagcgag gtatgtaggc ggtgctacag agttcttgaa gtggtggcct aactacggct
8521 acactagaag aacagtattt ggtatctgcg ctctgctgaa gccagttacc ttcggaaaaa
8581 gagttggtag ctcttgatcc ggcaaacaaa ccaccgctgg tagcggtttt tttgtttgca
8641 agcagcagat tacgcgcaga aaaaaaggat ctcaagaaga tcctttgatc ttttctacgg
8701 ggtctgacgc tcagtggaac gaaaactcac gttaagggat tttggtcatg agattatcaa
8761 aaaggatctt cacctagatc cttttaaatt aaaaatgaag ttttaaatca atctaaagta
8821 tatatgagta aacttggtct gacagttacc aatgcttaat cagtgaggca cctatctcag
8881 cgatctgtct atttcgttca tccatagttg cctgactccc cgtcgtgtag ataactacga
8941 tacgggaggg cttaccatct ggccccagtg ctgcaatgat accgcgagac ccacgctcac
9001 cggctccaga tttatcagca ataaaccagc cagccggaag ggccgagcgc agaagtggtc
9061 ctgcaacttt atccgcctcc atccagtcta ttaattgttg ccgggaagct agagtaagta
9121 gttcgccagt taatagtttg cgcaacgttg ttgccattgc tacaggcatc gtggtgtcac
9181 gctcgtcgtt tggtatggct tcattcagct ccggttccca acgatcaagg cgagttacat
9241 gatcccccat gttgtgcaaa aaagcggtta gctccttcgg tcctccgatc gttgtcagaa
9301 gtaagttggc cgcagtgtta tcactcatgg ttatggcagc actgcataat tctcttactg
9361 tcatgccatc cgtaagatgc ttttctgtga ctggtgagta ctcaaccaag tcattctgag
9421 aatagtgtat gcggcgaccg agttgctctt gcccggcgtc aatacgggat aataccgcgc
9481 cacatagcag aactttaaaa gtgctcatca ttggaaaacg ttcttcgggg cgaaaactct
9541 caaggatctt accgctgttg agatccagtt cgatgtaacc cactcgtgca cccaactgat
9601 cttcagcatc ttttactttc accagcgttt ctgggtgagc aaaaacagga aggcaaaatg
9661 ccgcaaaaaa gggaataagg gcgacacgga aatgttgaat actcatactc ttccttttc
9721 aatattattg aagcatttat cagggttatt gtctcatgag cggatacata tttgaatgta
9781 tttagaaaaa taaacaaata ggggttccgc gcacatttcc ccgaaaagtg ccacctgacg
9841 tc

LCV2_puro_CFTR_51_1217_gibson (SEQ ID NO: 35)

LOCUS Exported 14250 bp ds-DNA circular
DEFINITION synthetic circular DNA
KEYWORDS LCV2_puro_CFTR_51_1217_gibson
SOURCE synthetic DNA construct

ORGANISM recombinant plasmid
REFERENCE 1 (bases 1 to 14250)
FEATURES Location/Qualifiers
source 1..14250
/organism="recombinant plasmid"
/mol_type="other DNA"
misc_feature 1..33
/note="NLS"
misc_feature 34..57
/note="FLAG"
misc_feature 58..123
/note="P2A"
CDS 124..720
/note="Puro"
misc_binding 736..1324
/note="WPRE"
misc_feature 736..755
/note="mCherry_PCR_tail"
LTR 1395..1630
/note="3' LTR"
rep_origin 4079..4304
/note="ColE1"
misc_feature 4516..5322
/note="AmpR"
LTR 6472..6660
/note="5' LTR (R and U5 portions; U3 was replaced by the
CMV promoter)"
misc_feature 6711..6848
/note="Psi"
misc_feature 6768..6771
/note="SD; splice donor"
misc_feature 6815..7179
/note="gag"
misc_feature 7325..7566
/note="RRE"
misc_feature 8084..8201
/note="CPPT; central polypurine tract"
promoter 8252..8500
/note="Human U6"
misc_feature 8522..8607
/note="sgRNA scaffold"

misc_feature 8608..8613
/note="Linker"
promoter 8665..8920
/note="EFS-NS"
CDS 8944..10083
/codon_start=1
/note="ADARB1_Catalytic Domain" (SEQ ID NO: 36)
/translation="MLADAVSRLVLGKFGDLTDNFSSPHARRKVLAGVVMTTGTDVKDAKV
ISVSTGTKCINGEYMSDRGLALNDCHAEIISRRSLLRFLYTQLELYLNNKDDQKRSIFQ
KSERGGFRLKENVQFHLYISTSPCGDARIFSPHEPILEEPADRHPNRKARGQLRTKIES
GQGTIPVRSNASIQTWDGVLQGERLLTMSCSDKIARWNVVGIQGSLLSIFVEPIYFSSII
LGSLYHGDHLSRAMYQRISNIEDLPPLYTLNKPLLSGISNAEARQPGKAPNFSVNWT
VGDSAIEVINATTGKDELGRASRLCKHALYCRWMRVHGKVPSHLLRSKITKPNVYH
ESKLAAKEYQAAKARLFTAFIKAGLGAWVEKPTEQDQFSLTP"
misc_feature 8944..8946
/note="hSpCas9"
CDS 10084..10131
/codon_start=1
/note="XTEN"
/translation="SGSETPGTSESATPES" (SEQ ID NO: 37)
CDS 10132..14235
/codon_start=1
/product="catalytically dead mutant of the Cas9
endonuclease from the Streptococcus pyogenes Type II
CRISPR/Cas system"
/note="dCas9"
/note="RNA-guided DNA-binding protein that lacks
endonuclease activity due to the D10A mutation in the RuvC
catalytic domain and the H840A mutation in the HNH
catalytic domain" (SEQ ID NO: 38)
/translation="MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIG
ALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLV
EEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFR
GHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL
VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE
DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLA
RGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL
LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
KKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDR

EMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD
GFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV
VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKV
LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD
KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDF
QFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVR
KVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAY
SVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLP
KYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQK
QLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD"

ORIGIN (SEQ ID NO: 35)

1 acaaagaagg ctggacaggc taagaagaag aaagattaca aagacgatga cgataaggga
61 tccggcgcaa caaacttctc tctgctgaaa caagccggag atgtcgaaga gaatcctgga
121 ccgaccgagt acaagcccac ggtgcgcctc gccacccgcg acgacgtccc cagggccgta
181 cgcaccctcg ccgccgcgtt cgccgactac cccgccacgc gccacaccgt cgatccggac
241 cgccacatcg agcgggtcac cgagctgcaa gaactcttcc tcacgcgcgt cgggctcgac
301 atcggcaagg tgtgggtcgc ggacgacggc gccgcggtgg cggtctggac cacgccggag
361 agcgtcgaag cgggggcggt gttcgccgag atcggcccgc gcatggccga gttgagcggt
421 tcccggctgg ccgcgcagca acagatggaa ggcctcctgg cgccgcaccg gcccaaggag
481 cccgcgtggt tcctggccac cgtcggagtc tcgcccgacc accagggcaa gggtctgggc
541 agcgccgtcg tgctccccgg agtggaggcg gccgagcgcg ccggggtgcc cgccttcctg
601 gagacctccg cgccccgcaa cctccccttc tacgagcggc tcggcttcac cgtcaccgcc
661 gacgtcgagg tgcccgaagg accgcgcacc tggtgcatga cccgcaagcc cggtgcctga
721 acgcgttaag tcgacaatca acctctggat tacaaaattt gtgaaagatt gactggtatt
781 cttaactatg ttgctccttt tacgctatgt ggatacgctg ctttaatgcc tttgtatcat
841 gctattgctt cccgtatggc tttcattttc tcctccttgt ataaatcctg gttgctgtct
901 ctttatgagg agttgtggcc cgttgtcagg caacgtggcg tggtgtgcac tgtgtttgct
961 gacgcaaccc ccactggttg gggcattgcc accacctgtc agctcctttc cgggactttc
1021 gctttccccc tccctattgc cacggcggaa ctcatcgccg cctgccttgc ccgctgctgg
1081 acaggggctc ggctgttggg cactgacaat tccgtggtgt tgtcggggaa atcatcgtcc
1141 tttccttggc tgctcgcctg tgttgccacc tggattctgc gcgggacgtc cttctgctac
1201 gtcccttcgg ccctcaatcc agcggacctt ccttcccgcg gcctgctgcc ggctctgcgg
1261 cctcttccgc gtcttcgcct tcgccctcag acgagtcgga tctccctttg ggccgcctcc
1321 ccgcgtcgac tttaagacca atgacttaca aggcagctgt agatcttagc cacttttaa
1381 aagaaaaggg gggactggaa gggctaattc actcccaacg aagacaagat ctgctttttg
1441 cttgtactgg gtctctctgg ttagaccaga tctgagcctg ggagctctct ggctaactag
1501 ggaacccact gcttaagcct caataaagct tgccttgagt gcttcaagta gtgtgtgccc

1561 gtctgttgtg tgactctggt aactagagat ccctcagacc cttttagtca gtgtggaaaa
1621 tctctagcag ggcccgttta aacccgctga tcagcctcga ctgtgccttc tagttgccag
1681 ccatctgttg tttgcccctc ccccgtgcct tccttgaccc tggaaggtgc cactcccact
1741 gtcctttcct aataaaatga ggaaattgca tcgcattgtc tgagtaggtg tcattctatt
1801 ctgggggggtg gggtggggca ggacagcaag ggggaggatt gggaagacaa tagcaggcat
1861 gctggggatg cggtgggctc tatggcttct gaggcggaaa gaaccagctg gggctctagg
1921 gggtatcccc acgcgccctg tagcggcgca ttaagcgcgg cgggtgtggt ggttacgcgc
1981 agcgtgaccg ctacacttgc cagcgcccta gcgcccgctc ctttcgcttt cttcccttcc
2041 tttctcgcca cgttcgccgg ctttccccgt caagctctaa atcgggggct ccctttaggg
2101 ttccgattta gtgctttacg gcacctcgac cccaaaaaac ttgattaggg tgatggttca
2161 cgtagtgggc catcgccctg atagacggtt tttcgccctt tgacgttgga gtccacgttc
2221 tttaatagtg gactcttgtt ccaaactgga acaacactca accctatctc ggtctattct
2281 tttgatttat aagggatttt gccgatttcg gcctattggt taaaaaatga gctgatttaa
2341 caaaaattta acgcgaatta attctgtgga atgtgtgtca gttagggtgt ggaaagtccc
2401 caggctcccc agcaggcaga agtatgcaaa gcatgcatct caattagtca gcaaccaggt
2461 gtggaaagtc cccaggctcc ccagcaggca gaagtatgca aagcatgcat ctcaattagt
2521 cagcaaccat agtcccgccc ctaactccgc ccatcccgcc cctaactccg cccagttccg
2581 cccattctcc gccccatggc tgactaattt tttttattta tgcagaggcc gaggccgcct
2641 ctgcctctga gctattccag aagtagtgag gaggcttttt tggaggccta ggcttttgca
2701 aaaagctccc gggagcttgt atatccattt tcggatctga tcagcacgtg ttgacaatta
2761 atcatcggca tagtatatcg gcatagtata atacgacaag gtgaggaact aaaccatggc
2821 caagttgacc agtgccgttc cggtgctcac cgcgcgcgac gtcgccggag cggtcgagtt
2881 ctggaccgac cggctcgggt tctcccggga cttcgtggag gacgacttcg ccggtgtggt
2941 ccgggacgac gtgaccctgt tcatcagcgc ggtccaggac caggtggtgc cggacaacac
3001 cctggcctgg gtgtgggtgc gcggcctgga cgagctgtac gccgagtggt cggaggtcgt
3061 gtccacgaac ttccgggacg cctccgggcc ggccatgacc gagatcggcg agcagccgtg
3121 ggggcgggag ttcgccctgc gcgacccggc cggcaactgc gtgcacttcg tggccgagga
3181 gcaggactga cacgtgctac gagatttcga ttccaccgcc gccttctatg aaaggttggg
3241 cttcggaatc gttttccggg acgccggctg gatgatcctc cagcgcgggg atctcatgct
3301 ggagttcttc gcccacccca acttgtttat tgcagcttat aatggttaca aataaagcaa
3361 tagcatcaca aatttcacaa ataaagcatt tttttcactg cattctagtt gtggtttgtc
3421 caaactcatc aatgtatctt atcatgtctg tataccgtcg acctctagct agagcttggc
3481 gtaatcatgg tcatagctgt ttcctgtgtg aaattgttat ccgctcacaa ttccacacaa
3541 catacgagcc ggaagcataa agtgtaaagc ctggggtgcc taatgagtga gctaactcac
3601 attaattgcg ttgcgctcac tgcccgcttt ccagtcggga aacctgtcgt gccagctgca
3661 ttaatgaatc ggccaacgcg cggggagagg cggtttgcgt attgggcgct cttccgcttc
3721 ctcgctcact gactcgctgc gctcggtcgt tcggctgcgg cgagcggtat cagctcactc
3781 aaaggcggta atacggttat ccacagaatc aggggataac gcaggaaaga acatgtgagc
3841 aaaaggccag caaaaggcca ggaaccgtaa aaaggccgcg ttgctggcgt ttttccatag
3901 gctccgcccc cctgacgagc atcacaaaaa tcgacgctca agtcagaggt ggcgaaaccc
3961 gacaggacta taaagatacc aggcgtttcc ccctggaagc tccctcgtgc gctctcctgt

4021 tccgaccctg ccgcttaccg gatacctgtc cgcctttctc ccttcgggaa gcgtggcgct
4081 ttctcatagc tcacgctgta ggtatctcag ttcggtgtag gtcgttcgct ccaagctggg
4141 ctgtgtgcac gaacccccccg ttcagcccga ccgctgcgcc ttatccggta actatcgtct
4201 tgagtccaac ccggtaagac acgacttatc gccactggca gcagccactg gtaacaggat
4261 tagcagagcg aggtatgtag gcggtgctac agagttcttg aagtggtggc ctaactacgg
4321 ctacactaga agaacagtat ttggtatctg cgctctgctg aagccagtta ccttcggaaa
4381 aagagttggt agctcttgat ccggcaaaca aaccaccgct ggtagcggtg gttttttgt
4441 ttgcaagcag cagattacgc gcagaaaaaa aggatctcaa gaagatcctt tgatcttttc
4501 tacggggtct gacgctcagt ggaacgaaaa ctcacgttaa gggattttgg tcatgagatt
4561 atcaaaaagg atcttcacct agatcctttt aaattaaaaa tgaagtttta aatcaatcta
4621 aagtatatat gagtaaactt ggtctgacag ttaccaatgc ttaatcagtg aggcacctat
4681 ctcagcgatc tgtctatttc gttcatccat agttgcctga ctccccgtcg tgtagataac
4741 tacgatacgg gagggcttac catctggccc cagtgctgca atgataccgc gagacccacg
4801 ctcaccggct ccagatttat cagcaataaa ccagccagcc ggaagggccg agcgcagaag
4861 tggtcctgca actttatccg cctccatcca gtctattaat tgttgccggg aagctagagt
4921 aagtagttcg ccagttaata gtttgcgcaa cgttgttgcc attgctacag gcatcgtggt
4981 gtcacgctcg tcgtttggta tggcttcatt cagctccggt tcccaacgat caaggcgagt
5041 tacatgatcc cccatgttgt gcaaaaaagc ggttagctcc ttcggtcctc cgatcgttgt
5101 cagaagtaag ttggccgcag tgttatcact catggttatg gcagcactgc ataattctct
5161 tactgtcatg ccatccgtaa gatgcttttc tgtgactggt gagtactcaa ccaagtcatt
5221 ctgagaatag tgtatgcggc gaccgagttg ctcttgcccg gcgtcaatac gggataatac
5281 cgcgccacat agcagaactt taaaagtgct catcattgga aaacgttctt cggggcgaaa
5341 actctcaagg atcttaccgc tgttgagatc cagttcgatg taacccactc gtgcacccaa
5401 ctgatcttca gcatctttta ctttcaccag cgtttctggg tgagcaaaaa caggaaggca
5461 aaatgccgca aaaaagggaa taagggcgac acggaaatgt tgaatactca tactcttcct
5521 ttttcaatat tattgaagca tttatcaggg ttattgtctc atgagcggat acatatttga
5581 atgtatttag aaaaataaac aaataggggt tccgcgcaca tttccccgaa aagtgccacc
5641 tgacgtcgac ggatcgggag atctcccgat cccctatggt gcactctcag tacaatctgc
5701 tctgatgccg catagttaag ccagtatctg ctccctgctt gtgtgttgga ggtcgctgag
5761 tagtgcgcga gcaaaattta agctacaaca aggcaaggct tgaccgacaa ttgcatgaag
5821 aatctgctta gggttaggcg ttttgcgctg cttcgcgatg tacgggccag atatacgcgt
5881 tgacattgat tattgactag ttattaatag taatcaatta cggggtcatt agttcatagc
5941 ccatatatgg agttccgcgt tacataactt acggtaaatg gcccgcctgg ctgaccgccc
6001 aacgaccccc gcccattgac gtcaataatg acgtatgttc ccatagtaac gccaataggg
6061 actttccatt gacgtcaatg ggtggagtat ttacggtaaa ctgcccactt ggcagtacat
6121 caagtgtatc atatgccaag tacgcccccct attgacgtca atgacggtaa atggcccgcc
6181 tggcattatg cccagtacat gaccttatgg gactttccta cttggcagta catctacgta
6241 ttagtcatcg ctattaccat ggtgatgcgg ttttggcagt acatcaatgg gcgtggatag
6301 cggtttgact cacggggatt tccaagtctc caccccattg acgtcaatgg gagtttgttt
6361 tggcaccaaa atcaacggga ctttccaaaa tgtcgtaaca actccgcccc attgacgcaa
6421 atgggcggta ggcgtgtacg gtgggaggtc tatataagca gcgcgttttg cctgtactgg

6481 gtctctctgg ttagaccaga tctgagcctg ggagctctct ggctaactag ggaacccact
6541 gcttaagcct caataaagct tgccttgagt gcttcaagta gtgtgtgccc gtctgttgtg
6601 tgactctggt aactagagat ccctcagacc cttttagtca gtgtggaaaa tctctagcag
6661 tggcgcccga acagggactt gaaagcgaaa gggaaaccag aggagctctc tcgacgcagg
6721 actcggcttg ctgaagcgcg cacggcaaga ggcgaggggc ggcgactggt gagtacgcca
6781 aaaattttga ctagcggagg ctagaaggag agagatgggt gcgagagcgt cagtattaag
6841 cgggggagaa ttagatcgcg atgggaaaaa attcggttaa ggccaggggg aaagaaaaaa
6901 tataaattaa aacatatagt atgggcaagc agggagctag aacgattcgc agttaatcct
6961 ggcctgttag aaacatcaga aggctgtaga caaatactgg gacagctaca accatccctt
7021 cagacaggat cagaagaact tagatcatta tataatacag tagcaaccct ctattgtgtg
7081 catcaaagga tagagataaa agacaccaag gaagctttag acaagataga ggaagagcaa
7141 aacaaaagta agaccaccgc acagcaagcg gccgctgatc ttcagacctg gaggaggaga
7201 tatgagggac aattggagaa gtgaattata taaatataaa gtagtaaaaa ttgaaccatt
7261 aggagtagca cccaccaagg caaagagaag agtggtgcag agagaaaaaa gagcagtggg
7321 aataggagct ttgttccttg ggttcttggg agcagcagga agcactatgg gcgcagcgtc
7381 aatgacgctg acggtacagg ccagacaatt attgtctggt atagtgcagc agcagaacaa
7441 tttgctgagg gctattgagg cgcaacagca tctgttgcaa ctcacagtct ggggcatcaa
7501 gcagctccag gcaagaatcc tggctgtgga aagataccta aaggatcaac agctcctggg
7561 gatttggggt tgctctggaa aactcatttg caccactgct gtgccttgga atgctagttg
7621 gagtaataaa tctctggaac agatttggaa tcacacgacc tggatggagt gggacagaga
7681 aattaacaat tacacaagct taatacactc cttaattgaa gaatcgcaaa accagcaaga
7741 aaagaatgaa caagaattat tggaattaga taaatgggca agtttgtgga attggtttaa
7801 cataacaaat tggctgtggt atataaaatt attcataatg atagtaggag gcttggtagg
7861 tttaagaata gtttttgctg tactttctat agtgaataga gttaggcagg gatattcacc
7921 attatcgttt cagacccacc tcccaacccc gaggggaccc gacaggcccg aaggaataga
7981 agaagaaggt ggagagagag acagagacag atccattcga ttagtgaacg gatcggcact
8041 gcgtgcgcca attctgcaga caaatggcag tattcatcca caattttaaa agaaaagggg
8101 ggattggggg gtacagtgca ggggaaagaa tagtagacat aatagcaaca gacatacaaa
8161 ctaaagaatt acaaaaacaa attacaaaaa ttcaaaattt tcgggtttat tacagggaca
8221 gcagagatcc agtttggtta attaaggtac cgagggccta tttcccatga ttccttcata
8281 tttgcatata cgatacaagg ctgttagaga gataattaga attaatttga ctgtaaacac
8341 aaagatatta gtacaaaata cgtgacgtag aaagtaataa tttcttgggt agtttgcagt
8401 tttaaaatta tgttttaaaa tggactatca tatgcttacc gtaacttgaa agtatttcga
8461 tttcttggct ttatatatct tgtggaaagg acgaaacacc gttcataggg atccaagttt
8521 tgtttaagag ctatgctgga aacagcatag caagtttaaa taaggctagt ccgttatcaa
8581 cttgaaaaag tggcaccgag tcggtgcttc atttttcctc cactgttgca aagttttttt
8641 cctgcagccc gggaattcgc tagctaggtc ttgaaaggag tgggaattgg ctccggtgcc
8701 cgtcagtggg cagagcgcac atcgcccaca gtccccgaga agttgggggg aggggtcggc
8761 aattgatccg gtgcctagag aaggtggcgc ggggtaaact gggaaagtga tgtcgtgtac
8821 tggctccgcc tttttcccga gggtggggga gaaccgtata taagtgcagt agtcgccgtg
8881 aacgttcttt ttcgcaacgg gtttgccgcc agaacacagg accggttcta gagcgctgcc

8941 accatgttag ctgacgctgt ctcacgcctg gtcctgggta agtttggtga cctgaccgac
9001 aacttctcct cccctcacgc tcgcagaaaa gtgctggctg gagtcgtcat gacaacaggc
9061 acagatgtta aagatgccaa ggtgataagt gtttctacag gaacaaaatg tattaatggt
9121 gaatacatga gtgatcgtgg ccttgcatta aatgactgcc atgcagaaat aatatctcgg
9181 agatccttgc tcagatttct ttatacacaa cttgagcttt acttaaataa caaagatgat
9241 caaaaaagat ccatctttca gaaatcagag cgagggggggt ttaggctgaa ggagaatgtc
9301 cagtttcatc tgtacatcag cacctctccc tgtggagatg ccagaatctt ctcaccacat
9361 gagccaatcc tggaagaacc agcagataga cacccaaatc gtaaagcaag aggacagcta
9421 cggaccaaaa tagagtctgg tcaggggacg attccagtgc gctccaatgc gagcatccaa
9481 acgtgggacg gggtgctgca aggggagcgg ctgctcacca tgtcctgcag tgacaagatt
9541 gcacgctgga acgtggtggg catccaggga tccctgctca gcattttcgt ggagcccatt
9601 tacttctcga gcatcatcct gggcagcctt taccacgggg accacctttc cagggccatg
9661 taccagcgga tctccaacat agaggacctg ccacctctct acaccctcaa caagcctttg
9721 ctcagtggca tcagcaatgc agaagcacgg cagccaggga aggcccccaa cttcagtgtc
9781 aactggacgg taggcgactc cgctattgag gtcatcaacg ccacgactgg gaaggatgag
9841 ctgggccgcg cgtcccgcct gtgtaagcac gcgttgtact gtcgctggat gcgtgtgcac
9901 ggcaaggttc cctcccactt actacgctcc aagattacca agcccaacgt gtaccatgag
9961 tccaagctgg cggcaaagga gtaccaggcc gccaaggcgc gtctgttcac agccttcatc
10021 aaggcggggc tgggggcctg ggtggagaag cccaccgagc aggaccagtt ctcactcacg
10081 cccagtggaa gtgagacacc gggaacctca gagagcgcca cgccagaaag catggacaag
10141 aagtacagca tcggcctggc catcggcacc aactctgtgg gctgggccgt gatcaccgac
10201 gagtacaagg tgcccagcaa gaaattcaag gtgctgggca acaccgaccg gcacagcatc
10261 aagaagaacc tgatcggcgc cctgctgttc gacagcggag aaacagccga ggccacccgg
10321 ctgaagagaa ccgccagaag aagatacacc agacggaaga accggatctg ctatctgcaa
10381 gagatcttca gcaacgagat ggccaaggtg gacgacagct tcttccacag actggaagag
10441 tccttcctgg tggaagagga taagaagcac gagcggcacc ccatcttcgg caacatcgtg
10501 gacgaggtgg cctaccacga gaagtacccc accatctacc acctgagaaa gaaactggtg
10561 gacagcaccg acaaggccga cctgcggctg atctatctgg ccctggccca catgatcaag
10621 ttccggggcc acttcctgat cgagggcgac ctgaaccccg acaacagcga cgtggacaag
10681 ctgttcatcc agctggtgca gacctacaac cagctgttcg aggaaaaccc catcaacgcc
10741 agcggcgtgg acgccaaggc catcctgtct gccagactga gcaagagcag acggctggaa
10801 aatctgatcg cccagctgcc cggcgagaag aagaatggcc tgttcggcaa cctgattgcc
10861 ctgagcctgg gcctgacccc caacttcaag agcaacttcg acctggccga ggatgccaaa
10921 ctgcagctga gcaaggacac ctacgacgac gacctggaca acctgctggc ccagatcggc
10981 gaccagtacg ccgacctgtt tctggccgcc aagaacctgt ccgacgccat cctgctgagc
11041 gacatcctga gagtgaacac cgagatcacc aaggcccccc tgagcgcctc tatgatcaag
11101 agatacgacg agcaccacca ggacctgacc ctgctgaaag ctctcgtgcg gcagcagctg
11161 cctgagaagt acaaagagat tttcttcgac cagagcaaga acggctacgc cggctacatc
11221 gatggcggag ccagccagga agagttctac aagttcatca agcccatcct ggaaaagatg
11281 gacggcaccg aggaactgct cgtgaagctg aacagagagg acctgctgcg gaagcagcgg
11341 accttcgaca acggcagcat cccccaccag atccacctgg gagagctgca cgccattctg

11401 cggcggcagg aagattttta cccattcctg aaggacaacc gggaaaagat cgagaagatc
11461 ctgaccttcc gcatccccta ctacgtgggc cctctggcca ggggaaacag cagattcgcc
11521 tggatgacca gaaagagcga ggaaaccatc acccccctgga acttcgagga agtggtggac
11581 aagggcgcca gcgcccagag cttcatcgag cggatgacca acttcgataa gaacctgccc
11641 aacgagaagg tgctgcccaa gcacagcctg ctgtacgagt acttcaccgt gtacaacgag
11701 ctgaccaaag tgaaatacgt gaccgaggga atgagaaagc ccgccttcct gagcggcgag
11761 cagaaaaaag ccatcgtgga cctgctgttc aagaccaacc ggaaagtgac cgtgaagcag
11821 ctgaaagagg actacttcaa gaaaatcgag tgcttcgact ccgtggaaat ctccggcgtg
11881 gaagatcggt tcaacgcctc cctgggcaca taccacgatc tgctgaaaat tatcaaggac
11941 aaggacttcc tggacaatga ggaaaacgag gacattctgg aagatatcgt gctgaccctg
12001 acactgtttg aggacagaga gatgatcgag gaacggctga aaacctatgc ccacctgttc
12061 gacgacaaag tgatgaagca gctgaagcgg cggagataca ccggctgggg caggctgagc
12121 cggaagctga tcaacggcat ccgggacaag cagtccggca agacaatcct ggatttcctg
12181 aagtccgacg gcttcgccaa cagaaacttc atgcagctga tccacgacga cagcctgacc
12241 tttaaagagg acatccagaa agcccaggtg tccggccagg gcgatagcct gcacgagcac
12301 attgccaatc tggccggcag ccccgccatt aagaagggca tcctgcagac agtgaaggtg
12361 gtggacgagc tcgtgaaagt gatgggccgg cacaagcccg agaacatcgt gatcgaaatg
12421 gccagagaga accagaccac ccagaaggga cagaagaaca gccgcgagag aatgaagcgg
12481 atcgaagagg gcatcaaaga gctgggcagc cagatcctga aagaacaccc cgtggaaaac
12541 acccagctgc agaacgagaa gctgtacctg tactacctgc agaatgggcg ggatatgtac
12601 gtggaccagg aactggacat caaccggctg tccgactacg atgtggacgc tatcgtgcct
12661 cagagctttc tgaaggacga ctccatcgat aacaaagtgc tgactcggag cgacaagaac
12721 cggggcaaga gcgacaacgt gccctccgaa gaggtcgtga agaagatgaa gaactactgg
12781 cgccagctgc tgaatgccaa gctgattacc cagaggaagt tcgacaatct gaccaaggcc
12841 gagagaggcg gcctgagcga actggataag gccggcttca tcaagagaca gctggtggaa
12901 acccggcaga tcacaaagca cgtggcacag atcctggact cccggatgaa cactaagtac
12961 gacgagaacg acaaactgat ccgggaagtg aaagtgatca ccctgaagtc caagctggtg
13021 tccgatttcc ggaaggattt ccagttttac aaagtgcgcg agatcaacaa ctaccaccac
13081 gcccacgacg cctacctgaa cgccgtcgtg ggaaccgccc tgatcaaaaa gtaccctaag
13141 ctggaaagcg agttcgtgta cggcgactac aaggtgtacg acgtgcggaa gatgatcgcc
13201 aagagcgagc aggaaatcgg caaggctacc gccaagtact tcttctacag caacatcatg
13261 aacttttca agaccgagat taccctggcc aacggcgaga tccggaagcg gcctctgatc
13321 gagacaaacg gcgaaacagg cgagatcgtg tgggataagg gccgggactt tgccaccgtg
13381 cggaaagtgc tgtctatgcc ccaagtgaat atcgtgaaaa agaccgaggt gcagacaggc
13441 ggcttcagca aagagtctat cctgcccaag aggaacagcg acaagctgat cgccagaaag
13501 aaggactggg accctaagaa gtacggcggc ttcgacagcc ccaccgtggc ctattctgtg
13561 ctggtggtgg ccaaagtgga aaagggcaag tccaagaaac tgaagagtgt gaaagagctg
13621 ctggggatca ccatcatgga aagaagcagc ttcgagaaga atcccatcga ctttctggaa
13681 gccaagggct acaaagaagt gaaaaaggac ctgatcatca agctgcctaa gtactccctg
13741 ttcgagctgg aaaacggccg gaagagaatg ctggcctctg ccggcgaact gcagaaggga
13801 aacgaactgg ccctgccctc caaatatgtg aacttcctgt acctggccag ccactatgag

13861 aagctgaagg gctccccccga ggataatgag cagaaacagc tgtttgtgga acagcacaaa
13921 cactacctgg acgagatcat cgagcagatc agcgagttct ccaagagagt gatcctggcc
13981 gacgctaatc tggacaaggt gctgagcgcc tacaacaagc acagagacaa gcctatcaga
14041 gagcaggccg agaatatcat ccacctgttt accctgacca atctgggagc ccctgccgcc
14101 ttcaagtact ttgacaccac catcgaccgg aagaggtaca ccagcaccaa agaggtgctg
14161 gacgccaccc tgatccacca gagcatcacc ggcctgtacg agacacggat cgacctgtct
14221 cagctgggag gcgacaagcg acctgccgcc

AXCM_LCV2_puro_IDUA_No-spacer_gibson (SEQ ID NO: 39)

LOCUS Exported 14230 bp ds-DNA circular
DEFINITION synthetic circular DNA
KEYWORDS AXCM_LCV2_puro_IDUA_No-spacer_gibson
SOURCE synthetic DNA construct
ORGANISM synthetic DNA construct
REFERENCE 1 (bases 1 to 14230)
FEATURES Location/Qualifiers
source 1..14230
/organism="synthetic DNA construct"
/mol_type="other DNA"
LTR 828..1016
/note="5' LTR (R and U5 portions; U3 was replaced by the
CMV promoter)"
misc_feature 1067..1204
/note="Psi"
misc_feature 1124..1127
/note="SD; splice donor"
misc_feature 1171..1535
/note="gag"
misc_feature 1681..1922
/note="RRE"
misc_feature 2440..2557
/note="CPPT; central polypurine tract"
promoter 2608..2856
/note="Human U6"
misc_feature 2857..2942
/note="sgRNA scaffold"
misc_feature 2943..2948
/note="Linker"
promoter 3001..3256
/note="EFS-NS"

CDS 3280..4419
/codon_start=1
/note="ADARB1_Catalytic Domain" (SEQ ID NO: 40)
/translation="MLADAVSRLVLGKFGDLTDNFSSPHARRKVLAGVVMTTGTDVKDAKV
ISVSTGTKCINGEYMSDRGLALNDCHAEIISRRSLLRFLYTQLELYLNNKDDQKRSIFQ
KSERGGFRLKENVQFHLYISTSPCGDARIFSPHEPILEEPADRHPNRKARGQLRTKIES
GQGTIPVRSNASIQTWDGVLQGERLLTMSCSDKIARWNVVGIQGSLLSIFVEPIYFSSII
LGSLYHGDHLSRAMYQRISNIEDLPPLYTLNKPLLSGISNAEARQPGKAPNFSVNWT
VGDSAIEVINATTGKDELGRASRLCKHALYCRWMRVHGKVPSHLLRSKITKPNVYH
ESKLAAKEYQAAKARLFTAFIKAGLGAWVEKPTEQDQFSLTP"
misc_feature 3280..3282
/note="hSpCas9"
CDS 4420..4467
/codon_start=1
/note="XTEN"
/translation="SGSETPGTSESATPES" (SEQ ID NO: 41)
CDS 4468..8571
/codon_start=1
/product="catalytically dead mutant of the Cas9
endonuclease from the Streptococcus pyogenes Type II
CRISPR/Cas system"
/note="dCas9"
/note="RNA-guided DNA-binding protein that lacks
endonuclease activity due to the D10A mutation in the RuvC
catalytic domain and the H840A mutation in the HNH
catalytic domain" (SEQ ID NO: 42)

/translation="MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIG
ALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLV
EEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFR
GHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL
VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE
DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLA
RGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL
LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
KKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDR
EMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD
GFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV
VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKV
LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD
KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDF
QFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVR
KVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAY
SVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLP
KYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQK
QLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD"

misc_feature 8572..8619
/note="NLS"
CDS 8572
/codon_start=1
/product="catalytically dead mutant of the Cas9
endonuclease from the Streptococcus pyogenes Type II
CRISPR/Cas system"
/note="dCas9"
/note="RNA-guided DNA-binding protein that lacks
endonuclease activity due to the D10A mutation in the RuvC
catalytic domain and the H840A mutation in the HNH
catalytic domain"
/translation=""
misc_feature 8620..8643
/note="FLAG"
misc_feature 8644..8709
/note="P2A"
CDS 8710..9306
/note="Puro"
misc_binding 9322..9910
/note="WPRE"
LTR 9981..10216
/note="3' LTR"
rep_origin 12665..12890
/note="ColE1"
misc_feature 13102..13908
/note="AmpR"

ORIGIN (SEQ ID NO: 39)

1 gtcgacggat cgggagatct cccgatcccc tatggtgcac tctcagtaca atctgctctg
61 atgccgcata gttaagccag tatctgctcc ctgcttgtgt gttggaggtc gctgagtagt

121 gcgcgagcaa aatttaagct acaacaaggc aaggcttgac cgacaattgc atgaagaatc
181 tgcttagggt taggcgtttt gcgctgcttc gcgatgtacg ggccagatat acgcgttgac
241 attgattatt gactagttat taatagtaat caattacggg gtcattagtt catagcccat
301 atatggagtt ccgcgttaca taacttacgg taaatggccc gcctggctga ccgcccaacg
361 acccccgccc attgacgtca ataatgacgt atgttcccat agtaacgcca atagggactt
421 tccattgacg tcaatgggtg gagtatttac ggtaaactgc ccacttggca gtacatcaag
481 tgtatcatat gccaagtacg ccccctattg acgtcaatga cggtaaatgg cccgcctggc
541 attatgccca gtacatgacc ttatgggact ttcctacttg gcagtacatc tacgtattag
601 tcatcgctat taccatggtg atgcggtttt ggcagtacat caatgggcgt ggatagcggt
661 ttgactcacg gggatttcca agtctccacc ccattgacgt caatgggagt ttgttttggc
721 accaaaatca acgggacttt ccaaaatgtc gtaacaactc cgccccattg acgcaaatgg
781 gcggtaggcg tgtacggtgg gaggtctata taagcagcgc gttttgcctg tactgggtct
841 ctctggttag accagatctg agcctgggag ctctctggct aactagggaa cccactgctt
901 aagcctcaat aaagcttgcc ttgagtgctt caagtagtgt gtgcccgtct gttgtgtgac
961 tctggtaact agagatccct cagacccttt tagtcagtgt ggaaaatctc tagcagtggc
1021 gcccgaacag ggacttgaaa gcgaaaggga aaccagagga gctctctcga cgcaggactc
1081 ggcttgctga agcgcgcacg gcaagaggcg aggggcggcg actggtgagt acgccaaaaa
1141 ttttgactag cggaggctag aaggagagag atgggtgcga gagcgtcagt attaagcggg
1201 ggagaattag atcgcgatgg gaaaaaattc ggttaaggcc agggggaaag aaaaaatata
1261 aattaaaaca tatagtatgg gcaagcaggg agctagaacg attcgcagtt aatcctggcc
1321 tgttagaaac atcagaaggc tgtagacaaa tactgggaca gctacaacca tcccttcaga
1381 caggatcaga agaacttaga tcattatata atacagtagc aaccctctat tgtgtgcatc
1441 aaaggataga gataaaagac accaaggaag ctttagacaa gatagaggaa gagcaaaaca
1501 aaagtaagac caccgcacag caagcggccg ctgatcttca gacctggagg aggagatatg
1561 agggacaatt ggagaagtga attatataaa tataaagtag taaaaattga accattagga
1621 gtagcaccca ccaaggcaaa gagaagagtg gtgcagagag aaaaaagagc agtgggaata
1681 ggagctttgt tccttgggtt cttgggagca gcaggaagca ctatgggcgc agcgtcaatg
1741 acgctgacgg tacaggccag acaattattg tctggtatag tgcagcagca gaacaatttg
1801 ctgagggcta ttgaggcgca acagcatctg ttgcaactca cagtctgggg catcaagcag
1861 ctccaggcaa gaatcctggc tgtggaaaga tacctaaagg atcaacagct cctggggatt
1921 tggggttgct ctggaaaact catttgcacc actgctgtgc cttggaatgc tagttggagt
1981 aataaatctc tggaacagat ttggaatcac acgacctgga tggagtggga cagagaaatt
2041 aacaattaca caagcttaat acactcctta attgaagaat cgcaaaacca gcaagaaaag
2101 aatgaacaag aattattgga attagataaa tgggcaagtt tgtggaattg gtttaacata
2161 acaaattggc tgtggtatat aaaattattc ataatgatag taggaggctt ggtaggttta
2221 agaatagttt ttgctgtact ttctatagtg aatagagtta ggcagggata ttcaccatta
2281 tcgtttcaga cccacctccc aaccccgagg ggacccgaca ggcccgaagg aatagaagaa
2341 gaaggtggag agagagacag agacagatcc attcgattag tgaacggatc ggcactgcgt
2401 gcgccaattc tgcagacaaa tggcagtatt catccacaat tttaaaagaa aaggggggat
2461 tgggggggtac agtgcagggg aaagaatagt agacataata gcaacagaca tacaaactaa
2521 agaattacaa aaacaaatta caaaaattca aaattttcgg gtttattaca gggacagcag

2581 agatccagtt tggttaatta aggtaccgag ggcctatttc ccatgattcc ttcatatttg
2641 catatacgat acaaggctgt tagagagata attagaatta atttgactgt aaacacaaag
2701 atattagtac aaaatacgtg acgtagaaag taataatttc ttgggtagtt tgcagtttta
2761 aaattatgtt ttaaaatgga ctatcatatg cttaccgtaa cttgaaagta tttcgatttc
2821 ttggctttat atatcttgtg gaaaggacga aacaccgttt aagagctatg ctggaaacag
2881 catagcaagt ttaaataagg ctagtccgtt atcaacttga aaaagtggca ccgagtcggt
2941 gcttcattac ttcggcccag agctgctcct ttttttcctg cagcccggga attcgctagc
3001 taggtcttga aaggagtggg aattggctcc ggtgcccgtc agtgggcaga gcgcacatcg
3061 cccacagtcc ccgagaagtt ggggggaggg gtcggcaatt gatccggtgc ctagagaagg
3121 tggcgcgggg taaactggga aagtgatgtc gtgtactggc tccgcctttt tcccgagggt
3181 gggggagaac cgtatataag tgcagtagtc gccgtgaacg ttcttttttcg caacgggttt
3241 gccgccagaa cacaggaccg gttctagagc gctgccacca tgttagctga cgctgtctca
3301 cgcctggtcc tgggtaagtt tggtgacctg accgacaact tctcctcccc tcacgctcgc
3361 agaaaagtgc tggctggagt cgtcatgaca acaggcacag atgttaaaga tgccaaggtg
3421 ataagtgttt ctacaggaac aaaatgtatt aatggtgaat acatgagtga tcgtggcctt
3481 gcattaaatg actgccatgc agaaataata tctcggagat ccttgctcag atttctttat
3541 acacaacttg agctttactt aaataacaaa gatgatcaaa aaagatccat ctttcagaaa
3601 tcagagcgag gggggtttag gctgaaggag aatgtccagt ttcatctgta catcagcacc
3661 tctccctgtg gagatgccag aatcttctca ccacatgagc caatcctgga agaaccagca
3721 gatagacacc caaatcgtaa agcaagagga cagctacgga ccaaaataga gtctggtcag
3781 gggacgattc cagtgcgctc caatgcgagc atccaaacgt gggacggggt gctgcaaggg
3841 gagcggctgc tcaccatgtc ctgcagtgac aagattgcac gctggaacgt ggtgggcatc
3901 cagggatccc tgctcagcat tttcgtggag cccatttact tctcgagcat catcctgggc
3961 agcctttacc acggggacca cctttccagg gccatgtacc agcggatctc caacatagag
4021 gacctgccac ctctctacac cctcaacaag cctttgctca gtggcatcag caatgcagaa
4081 gcacggcagc cagggaaggc ccccaacttc agtgtcaact ggacggtagg cgactccgct
4141 attgaggtca tcaacgccac gactgggaag gatgagctgg gccgcgcgtc ccgcctgtgt
4201 aagcacgcgt tgtactgtcg ctggatgcgt gtgcacggca aggttccctc ccacttacta
4261 cgctccaaga ttaccaagcc caacgtgtac catgagtcca agctggcggc aaaggagtac
4321 caggccgcca aggcgcgtct gttcacagcc ttcatcaagg cggggctggg ggcctgggtg
4381 gagaagccca ccgagcagga ccagttctca ctcacgccca gtggaagtga gacaccggga
4441 acctcagaga gcgccacgcc agaaagcatg gacaagaagt acagcatcgg cctggccatc
4501 ggcaccaact ctgtgggctg ggccgtgatc accgacgagt acaaggtgcc cagcaagaaa
4561 ttcaaggtgc tgggcaacac cgaccggcac agcatcaaga agaacctgat cggcgccctg
4621 ctgttcgaca gcggagaaac agccgaggcc acccggctga agagaaccgc cagaagaaga
4681 tacaccagac ggaagaaccg gatctgctat ctgcaagaga tcttcagcaa cgagatggcc
4741 aaggtggacg acagcttctt ccacagactg gaagagtcct tcctggtgga agaggataag
4801 aagcacgagc ggcaccccat cttcggcaac atcgtggacg aggtggccta ccacgagaag
4861 taccccacca tctaccacct gagaaagaaa ctggtggaca gcaccgacaa ggccgacctg
4921 cggctgatct atctggccct ggcccacatg atcaagttcc ggggccactt cctgatcgag
4981 ggcgacctga accccgacaa cagcgacgtg gacaagctgt tcatccagct ggtgcagacc

5041 tacaaccagc tgttcgagga aaaccccatc aacgccagcg gcgtggacgc caaggccatc
5101 ctgtctgcca gactgagcaa gagcagacgg ctggaaaatc tgatcgccca gctgcccggc
5161 gagaagaaga atggcctgtt cggcaacctg attgccctga gcctgggcct gacccccaac
5221 ttcaagagca acttcgacct ggccgaggat gccaaactgc agctgagcaa ggacacctac
5281 gacgacgacc tggacaacct gctggcccag atcggcgacc agtacgccga cctgtttctg
5341 gccgccaaga acctgtccga cgccatcctg ctgagcgaca tcctgagagt gaacaccgag
5401 atcaccaagg ccccccctgag cgcctctatg atcaagagat acgacgagca ccaccaggac
5461 ctgaccctgc tgaaagctct cgtgcggcag cagctgcctg agaagtacaa agagattttc
5521 ttcgaccaga gcaagaacgg ctacgccggc tacatcgatg gcggagccag ccaggaagag
5581 ttctacaagt tcatcaagcc catcctggaa aagatggacg gcaccgagga actgctcgtg
5641 aagctgaaca gagaggacct gctgcggaag cagcggacct tcgacaacgg cagcatcccc
5701 caccagatcc acctgggaga gctgcacgcc attctgcggc ggcaggaaga tttttaccca
5761 ttcctgaagg acaaccggga aaagatcgag aagatcctga ccttccgcat cccctactac
5821 gtgggccctc tggccagggg aaacagcaga ttcgcctgga tgaccagaaa gagcgaggaa
5881 accatcaccc cctggaactt cgaggaagtg gtggacaagg gcgccagcgc ccagagcttc
5941 atcgagcgga tgaccaactt cgataagaac ctgcccaacg agaaggtgct gcccaagcac
6001 agcctgctgt acgagtactt caccgtgtac aacgagctga ccaaagtgaa atacgtgacc
6061 gagggaatga gaaagcccgc cttcctgagc ggcgagcaga aaaaagccat cgtggacctg
6121 ctgttcaaga ccaaccggaa agtgaccgtg aagcagctga aagaggacta cttcaagaaa
6181 atcgagtgct tcgactccgt ggaaatctcc ggcgtggaag atcggttcaa cgcctccctg
6241 ggcacatacc acgatctgct gaaaattatc aaggacaagg acttcctgga caatgaggaa
6301 aacgaggaca ttctggaaga tatcgtgctg accctgacac tgtttgagga cagagagatg
6361 atcgaggaac ggctgaaaac ctatgcccac ctgttcgacg acaaagtgat gaagcagctg
6421 aagcggcgga gatacaccgg ctggggcagg ctgagccgga agctgatcaa cggcatccgg
6481 gacaagcagt ccggcaagac aatcctggat ttcctgaagt ccgacggctt cgccaacaga
6541 aacttcatgc agctgatcca cgacgacagc ctgaccttta aagaggacat ccagaaagcc
6601 caggtgtccg gccagggcga tagcctgcac gagcacattg ccaatctggc cggcagcccc
6661 gccattaaga agggcatcct gcagacagtg aaggtggtgg acgagctcgt gaaagtgatg
6721 ggccggcaca agcccgagaa catcgtgatc gaaatggcca gagagaacca gaccacccag
6781 aagggacaga agaacagccg cgagagaatg aagcggatcg aagagggcat caaagagctg
6841 ggcagccaga tcctgaaaga acaccccgtg gaaaacaccc agctgcagaa cgagaagctg
6901 tacctgtact acctgcagaa tgggcgggat atgtacgtgg accaggaact ggacatcaac
6961 cggctgtccg actacgatgt ggacgctatc gtgcctcaga gctttctgaa ggacgactcc
7021 atcgataaca aagtgctgac tcggagcgac aagaaccggg gcaagagcga caacgtgccc
7081 tccgaagagg tcgtgaagaa gatgaagaac tactggcgcc agctgctgaa tgccaagctg
7141 attacccaga ggaagttcga caatctgacc aaggccgaga gaggcggcct gagcgaactg
7201 gataaggccg gcttcatcaa gagacagctg gtggaaaccc ggcagatcac aaagcacgtg
7261 gcacagatcc tggactcccg gatgaacact aagtacgacg agaacgacaa actgatccgg
7321 gaagtgaaag tgatcaccct gaagtccaag ctggtgtccg atttccggaa ggatttccag
7381 ttttacaaag tgcgcgagat caacaactac caccacgccc acgacgccta cctgaacgcc
7441 gtcgtgggaa ccgccctgat caaaaagtac cctaagctgg aaagcgagtt cgtgtacggc

7501 gactacaagg tgtacgacgt gcggaagatg atcgccaaga gcgagcagga aatcggcaag
7561 gctaccgcca agtacttctt ctacagcaac atcatgaact ttttcaagac cgagattacc
7621 ctggccaacg gcgagatccg gaagcggcct ctgatcgaga caaacggcga aacaggcgag
7681 atcgtgtggg ataagggccg ggactttgcc accgtgcgga aagtgctgtc tatgccccaa
7741 gtgaatatcg tgaaaaagac cgaggtgcag acaggcggct tcagcaaaga gtctatcctg
7801 cccaagagga acagcgacaa gctgatcgcc agaaagaagg actgggaccc taagaagtac
7861 ggcggcttcg acagccccac cgtggcctat tctgtgctgg tggtggccaa agtggaaaag
7921 ggcaagtcca agaaactgaa gagtgtgaaa gagctgctgg ggatcaccat catggaaaga
7981 agcagcttcg agaagaatcc catcgacttt ctggaagcca agggctacaa agaagtgaaa
8041 aaggacctga tcatcaagct gcctaagtac tccctgttcg agctggaaaa cggccggaag
8101 agaatgctgg cctctgccgg cgaactgcag aagggaaacg aactggccct gccctccaaa
8161 tatgtgaact tcctgtacct ggccagccac tatgagaagc tgaagggctc ccccgaggat
8221 aatgagcaga aacagctgtt tgtggaacag cacaaacact acctggacga gatcatcgag
8281 cagatcagcg agttctccaa gagagtgatc ctggccgacg ctaatctgga caaggtgctg
8341 agcgcctaca acaagcacag agacaagcct atcagagagc aggccgagaa tatcatccac
8401 ctgtttaccc tgaccaatct gggagcccct gccgccttca agtactttga caccaccatc
8461 gaccggaaga ggtacaccag caccaaagag gtgctggacg ccaccctgat ccaccagagc
8521 atcaccggcc tgtacgagac acggatcgac ctgtctcagc tgggaggcga caagcgacct
8581 gccgccacaa agaaggctgg acaggctaag aagaagaaag attacaaaga cgatgacgat
8641 aagggatccg gcgcaacaaa cttctctctg ctgaaacaag ccggagatgt cgaagagaat
8701 cctggaccga ccgagtacaa gcccacggtg cgcctcgcca cccgcgacga cgtccccagg
8761 gccgtacgca ccctcgccgc cgcgttcgcc gactaccccg ccacgcgcca caccgtcgat
8821 ccggaccgcc acatcgagcg ggtcaccgag ctgcaagaac tcttcctcac gcgcgtcggg
8881 ctcgacatcg gcaaggtgtg ggtcgcggac gacggcgccg cggtggcggt ctggaccacg
8941 ccggagagcg tcgaagcggg ggcggtgttc gccgagatcg gcccgcgcat ggccgagttg
9001 agcggttccc ggctggccgc gcagcaacag atggaaggcc tcctggcgcc gcaccggccc
9061 aaggagcccg cgtggttcct ggccaccgtc ggagtctcgc ccgaccacca gggcaagggt
9121 ctgggcagcg ccgtcgtgct ccccggagtg gaggcggccg agcgcgccgg ggtgcccgcc
9181 ttcctggaga cctccgcgcc ccgcaacctc cccttctacg agcggctcgg cttcaccgtc
9241 accgccgacg tcgaggtgcc cgaaggaccg cgcacctggt gcatgacccg caagcccggt
9301 gcctgaacgc gttaagtcga caatcaacct ctggattaca aaatttgtga aagattgact
9361 ggtattctta actatgttgc tccttttacg ctatgtggat acgctgcttt aatgcctttg
9421 tatcatgcta ttgcttcccg tatggctttc attttctcct ccttgtataa atcctggttg
9481 ctgtctcttt atgaggagtt gtggcccgtt gtcaggcaac gtggcgtggt gtgcactgtg
9541 tttgctgacg caacccccac tggttggggc attgccacca cctgtcagct cctttccggg
9601 actttcgctt tccccctccc tattgccacg gcggaactca tcgccgcctg ccttgcccgc
9661 tgctggacag gggctcggct gttgggcact gacaattccg tggtgttgtc ggggaaatca
9721 tcgtcctttc cttggctgct cgcctgtgtt gccacctgga ttctgcgcgg gacgtccttc
9781 tgctacgtcc cttcggccct caatccagcg gaccttcctt cccgcggcct gctgccggct
9841 ctgcggcctc ttccgcgtct tcgccttcgc cctcagacga gtcggatctc cctttgggcc
9901 gcctccccgc gtcgacttta agaccaatga cttacaaggc agctgtagat cttagccact

9961 ttttaaaaga aaaggggga ctggaagggc taattcactc ccaacgaaga caagatctgc
10021 tttttgcttg tactgggtct ctctggttag accagatctg agcctgggag ctctctggct
10081 aactagggaa cccactgctt aagcctcaat aaagcttgcc ttgagtgctt caagtagtgt
10141 gtgcccgtct gttgtgtgac tctggtaact agagatccct cagacccttt tagtcagtgt
10201 ggaaaatctc tagcagggcc cgtttaaacc cgctgatcag cctcgactgt gccttctagt
10261 tgccagccat ctgttgtttg cccctccccc gtgccttcct tgaccctgga aggtgccact
10321 cccactgtcc tttcctaata aaatgaggaa attgcatcgc attgtctgag taggtgtcat
10381 tctattctgg ggggtggggt ggggcaggac agcaaggggg aggattggga agacaatagc
10441 aggcatgctg gggatgcggt gggctctatg gcttctgagg cggaaagaac cagctggggc
10501 tctagggggt atccccacgc gccctgtagc ggcgcattaa gcgcggcggg tgtggtggtt
10561 acgcgcagcg tgaccgctac acttgccagc gccctagcgc ccgctccttt cgctttcttc
10621 ccttcctttc tcgccacgtt cgccggcttt ccccgtcaag ctctaaatcg ggggctccct
10681 ttagggttcc gatttagtgc tttacggcac ctcgacccca aaaaacttga ttagggtgat
10741 ggttcacgta gtgggccatc gccctgatag acggtttttc gcccctttgac gttggagtcc
10801 acgttcttta atagtggact cttgttccaa actggaacaa cactcaaccc tatctcggtc
10861 tattcttttg atttataagg gattttgccg atttcggcct attggttaaa aaatgagctg
10921 atttaacaaa aatttaacgc gaattaattc tgtggaatgt gtgtcagtta gggtgtggaa
10981 agtccccagg ctccccagca ggcagaagta tgcaaagcat gcatctcaat tagtcagcaa
11041 ccaggtgtgg aaagtcccca ggctccccag caggcagaag tatgcaaagc atgcatctca
11101 attagtcagc aaccatagtc ccgcccctaa ctccgcccat cccgccccta actccgccca
11161 gttccgccca ttctccgccc catggctgac taatttttt tatttatgca gaggccgagg
11221 ccgcctctgc ctctgagcta ttccagaagt agtgaggagg ctttttgga ggcctaggct
11281 tttgcaaaaa gctcccggga gcttgtatat ccattttcgg atctgatcag cacgtgttga
11341 caattaatca tcggcatagt atatcggcat agtataatac gacaaggtga ggaactaaac
11401 catggccaag ttgaccagtg ccgttccggt gctcaccgcg cgcgacgtcg ccggagcggt
11461 cgagttctgg accgaccggc tcgggttctc ccgggacttc gtggaggacg acttcgccgg
11521 tgtggtccgg gacgacgtga ccctgttcat cagcgcggtc caggaccagg tggtgccgga
11581 caacaccctg gcctgggtgt gggtgcgcgg cctggacgag ctgtacgccg agtggtcgga
11641 ggtcgtgtcc acgaacttcc gggacgcctc cgggccggcc atgaccgaga tcggcgagca
11701 gccgtggggg cgggagttcg ccctgcgcga cccggccggc aactgcgtgc acttcgtggc
11761 cgaggagcag gactgacacg tgctacgaga tttcgattcc accgccgcct tctatgaaag
11821 gttgggcttc ggaatcgttt tccgggacgc cggctggatg atcctccagc gcggggatct
11881 catgctggag ttcttcgccc accccaactt gtttattgca gcttataatg gttacaaata
11941 aagcaatagc atcacaaatt tcacaaataa agcatttttt tcactgcatt ctagttgtgg
12001 tttgtccaaa ctcatcaatg tatcttatca tgtctgtata ccgtcgacct ctagctagag
12061 cttggcgtaa tcatggtcat agctgtttcc tgtgtgaaat tgttatccgc tcacaattcc
12121 acacaacata cgagccggaa gcataaagtg taaagcctgg ggtgcctaat gagtgagcta
12181 actcacatta attgcgttgc gctcactgcc cgctttccag tcgggaaacc tgtcgtgcca
12241 gctgcattaa tgaatcggcc aacgcgcggg gagaggcggt ttgcgtattg ggcgctcttc
12301 cgcttcctcg ctcactgact cgctgcgctc ggtcgttcgg ctgcggcgag cggtatcagc
12361 tcactcaaag gcggtaatac ggttatccac agaatcaggg gataacgcag gaaagaacat

12421 gtgagcaaaa ggccagcaaa aggccaggaa ccgtaaaaag gccgcgttgc tggcgttttt
12481 ccataggctc cgccccccctg acgagcatca caaaaatcga cgctcaagtc agaggtggcg
12541 aaacccgaca ggactataaa gataccaggc gtttcccct ggaagctccc tcgtgcgctc
12601 tcctgttccg accctgccgc ttaccggata cctgtccgcc tttctccctt cgggaagcgt
12661 ggcgctttct catagctcac gctgtaggta tctcagttcg gtgtaggtcg ttcgctccaa
12721 gctgggctgt gtgcacgaac cccccgttca gcccgaccgc tgcgccttat ccggtaacta
12781 tcgtcttgag tccaacccgg taagacacga cttatcgcca ctggcagcag ccactggtaa
12841 caggattagc agagcgaggt atgtaggcgg tgctacagag ttcttgaagt ggtggcctaa
12901 ctacggctac actagaagaa cagtatttgg tatctgcgct ctgctgaagc cagttacctt
12961 cggaaaaaga gttggtagct cttgatccgg caaacaaacc accgctggta gcggtggttt
13021 ttttgtttgc aagcagcaga ttacgcgcag aaaaaaagga tctcaagaag atcctttgat
13081 cttttctacg gggtctgacg ctcagtggaa cgaaaactca cgttaaggga ttttggtcat
13141 gagattatca aaaaggatct tcacctagat ccttttaaat taaaaatgaa gttttaaatc
13201 aatctaaagt atatatgagt aaacttggtc tgacagttac caatgcttaa tcagtgaggc
13261 acctatctca gcgatctgtc tatttcgttc atccatagtt gcctgactcc ccgtcgtgta
13321 gataactacg atacgggagg gcttaccatc tggccccagt gctgcaatga taccgcgaga
13381 cccacgctca ccggctccag atttatcagc aataaaccag ccagccggaa gggccgagcg
13441 cagaagtggt cctgcaactt tatccgcctc catccagtct attaattgtt gccgggaagc
13501 tagagtaagt agttcgccag ttaatagttt gcgcaacgtt gttgccattg ctacaggcat
13561 cgtggtgtca cgctcgtcgt ttggtatggc ttcattcagc tccggttccc aacgatcaag
13621 gcgagttaca tgatccccca tgttgtgcaa aaaagcggtt agctccttcg gtcctccgat
13681 cgttgtcaga agtaagttgg ccgcagtgtt atcactcatg gttatggcag cactgcataa
13741 ttctcttact gtcatgccat ccgtaagatg cttttctgtg actggtgagt actcaaccaa
13801 gtcattctga gaatagtgta tgcggcgacc gagttgctct tgcccggcgt caatacggga
13861 taataccgcg ccacatagca gaactttaaa agtgctcatc attggaaaac gttcttcggg
13921 gcgaaaactc tcaaggatct taccgctgtt gagatccagt tcgatgtaac ccactcgtgc
13981 acccaactga tcttcagcat cttttacttt caccagcgtt tctgggtgag caaaaacagg
14041 aaggcaaaat gccgcaaaaa agggaataag ggcgacacgg aaatgttgaa tactcatact
14101 cttccttttt caatattatt gaagcattta tcagggttat tgtctcatga gcggatacat
14161 atttgaatgt atttagaaaa ataaacaaat aggggttccg cgcacatttc cccgaaaagt
14221 gccacctgac

WHAT IS CLAIMED IS:

1. A recombinant expression system for CRISPR/Cas-directed RNA editing of a target RNA comprising:

(A) a nucleic acid sequence encoding a CRISPR/Cas RNA editing fusion protein comprising a nuclease-dead CRISPR associated endonuclease (dCas) fused to a catalytically active deaminase domain of Adenosine Deaminase acting on RNA 2 (ADAR2); and

(B) a nucleic acid sequence encoding an extended single guide RNA (esgRNA) comprising: (i) a short extension sequence of homology to the target RNA comprising a mismatch for a target adenosine, and (ii) a dCas9 scaffold binding sequence.

2. The recombinant expression system of claim 1, wherein the esgRNA further comprises (iii) a spacer sequence comprising a region of homology to the target RNA.

3. The recombinant expression system of claim 1, wherein (A) and (B) are comprised within the same vector or comprised within different vectors.

4. The recombinant expression system of claim 1, wherein the catalytically active deaminase domain of ADAR2 is (1) a wildtype catalytically active deaminase domain of human ADAR2 or (2) a mutant human catalytically active deaminase domain of ADAR2 with increased catalytic activity compared to the wildtype human ADAR2.

5. The recombinant expression system of claim 4, wherein the mutant human catalytically active deaminase domain of ADAR2 comprises a E488Q mutation.

6. The recombinant expression system of claim 1, wherein the dCas9 N-terminal domain is fused to the C-terminus of the catalytically active deaminase domain of ADAR2.

7. The recombinant expression system of any one of the preceding claims, wherein the dCas9 is fused to the catalytically active deaminase domain of ADAR2 via a linker.

8. The recombinant expression system of claim 7, wherein the linker is a semi-flexible XTEN peptide linker.

9. The recombinant expression system of any one of the previous claims, wherein the short extension sequence of the esgRNA is a 3’ extension sequence.

10. The recombinant expression system of any one of the previous claims, wherein the short extension sequence of the esgRNA comprises a region of homology capable of near-perfect RNA-RNA base pairing with the target sequence.

11. The recombinant expression system of any one of the preceding claims, wherein the short extension sequence of the esgRNA further comprises a second mismatch for an adenosine within the target RNA.

12. The recombinant expression system of claim 11, wherein the short extension sequence of the esgRNA further comprises a third mismatch for an adenosine within the target RNA and optionally a fourth mismatch for an adenosine within the target RNA.

13. The recombinant expression system of any one of the previous claims, wherein the short extension sequence of the esgRNA is 15 nucleotides to 60 nucleotides in length.

14. The recombinant expression system of any one of the previous claims, wherein the esgRNA further comprises a marker sequence.

15. The recombinant expression system of any one of the previous claims, wherein the esgRNA further comprises a RNA polymerase III promoter sequence.

16. The recombinant expression system of claim 15, wherein the RNA polymerase III promoter sequence is a U6 promoter sequence.

17. The recombinant expression system of any one of the previous claims, wherein the esgRNA comprises a linker sequence between a spacer sequence and a scaffold sequence.

18. The recombinant expression system of claim 2, wherein the sequences of the esgRNA (i), (ii), and (iii) are situated 3’ to 5’ in the esgRNA.

19. The recombinant expression system of any one of the previous claims, further comprising a nucleic acid encoding a PAM sequence or complement thereof.

20. The recombinant expression system of claim 3, wherein the vector is a viral vector.

21. The recombinant expression system of claim 20, wherein the viral vector is an adeno-associated viral vector (AAV), lentiviral vector, or an adenoviral vector.

22. A cell comprising any one of the recombinant expression systems of claims 1 to 21.

23. A method of selective RNA editing comprising administering any one of the recombinant expression systems of claims 1 to 21 to a cell, wherein the method is *in vitro*.

24. The method of claim 23, further comprising administering an antisense synthetic oligonucleotide compound comprising alternating 2'OMe RNA and DNA bases (PAMmer).

25. A method of characterizing the effects of directed cellular RNA editing on processing and dynamics comprising administering any one of the recombinant expression systems of claims 1 to 21 to a sample and determining its effects.

26. The method of claim 25, wherein the sample is derived from a subject.

27. Use of any one of the recombinant expression systems of claims 1 to 21, or the cell of claim 22, in the manufacture of a medicament for treating a disease selected from the group consisting of Hurler's syndrome, Cystic fibrosis, muscular dystrophy, spinal cord injury, hearing loss (through noise overexposure or ototoxicity), multiple sclerosis, Alzheimer's disease, amyotrophic lateral sclerosis (ALS), Parkinson's disease, over-rapid benzodiazepine withdrawal, and Huntington's disease in a subject, wherein the medicament is to be administered to a subject or a sample isolated from a subject.

28. The use of claim 27, further comprising correction of a G to A mutation in a target RNA.

29. A kit comprising:

(A) one or more of:

(i) a recombinant expression system according to any one of claims 1 to 21;

(ii) a viral particle comprising the recombinant expression system according to any one of claims 1 to 21; and/or

(iii) a cell comprising the recombinant expression system according to any one of claims 1 to 21; and

(B) instructions for use.

30. The kit of claim 29, wherein the instructions for use are according to a method comprising administering the one or more of part (A) to a cell, a sample isolated from a subject, or to a subject.