ENGINEERED CAS9 SYSTEMS FOR EUKARYOTIC GENOME MODIFICATION

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application Serial No. 62/631,304, filed February 15, 2018, and U.S. Provisional Application Serial No. 62/720,525, filed August 21, 2018, the disclosure of each of which is hereby incorporated by reference in its entirety.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on February 12, 2019, is named P18_023PCT_SL.txt and is 370,305 bytes in size.

FIELD

[0003] The present disclosure relates to engineered Cas9 systems, nucleic acids encoding said systems, and methods of using said systems for genome modification.

BACKGROUND

[0004] The recent development of the bacterial class 2 Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated (Cas) CRISPR/Cas systems as genome editing tools has provided unprecedented ease and simplicity to engineer site-specific endonucleases for eukaryotic genome modification. However, because each CRISPR/Cas system requires a specific protospacer adjacent motif (PAM) for target DNA binding, each system is limited to certain genomic sites. Although the currently most widespread adopted *Streptococcus pyogenes* Cas9 (SpyCas9) uses a frequently occurring PAM (5'-NGG-3') for targeting, it is still excluded from many genomic sites lacking such a motif, since eukaryotic genomes, especially those of mammals and plants, are highly complex and

heterogeneous in DNA sequence. Moreover, precision gene editing using homology-directed repair (HDR) or base editors such as dCas9/cytidine deaminase and dCas9/adenosine deaminase often requires a precise DNA binding position, even at the single base pair resolution, to achieve an optimal editing outcome. Therefore, there is a need to develop new CRISPR/Cas systems that use novel PAMs for targeting to increase genome coverage density.

SUMMARY

[0005] Among the various aspects of the present disclosure include engineered Cas9 systems comprising engineered Cas9 proteins and engineered guide RNAs, wherein each engineered guide RNA is designed to complex with an engineered Cas9 protein and the engineered guide RNA comprises a 5' guide sequence designed to hybridize with a target sequence in a double-stranded sequence, wherein the target sequence is 5' to a protospacer adjacent motif (PAM) and the PAM has a sequence as listed in Table A.

[0006] Another aspect of the present disclosure encompasses a plurality of nucleic acids encoding said engineered Cas9 systems and at least one vector comprising the plurality of said nucleic acids.

[0007] A further aspect includes eukaryotic cells comprising at least one engineered Cas9 system and/or at least one nucleic acid encoding said engineered Cas9 system.

[0008] Still another aspect of the present disclosure encompasses methods for modifying chromosomal sequences in eukaryotic cells. The methods comprise introducing into the eukaryotic cell at least one engineered Cas9 system comprising an engineered Cas9 protein and an engineered guide RNA and/or at least one nucleic acid encoding said engineered Cas9 system and, optionally, at least one donor polynucleotide, wherein the at least one engineered guide RNA guides the at least one engineered Cas9 protein to the target site in the chromosomal sequence such that modification of the chromosomal sequence occurs.

[0009] Other aspects and features of the disclosure are detailed bellow.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] FIG. **1** shows the WebLogo analysis of protospacer adjacent motifs (PAM) required for *in vitro* target DNA cleavage by Cas9 orthologs. Numbers on the horizontal axis indicate the position of the nucleotide in the PAM sequence.

[0011] FIG. **2A** presents the cleavage efficiency (as the percent of indels) of McaCas9, McaCas9-HN1HB1 fusion (*i.e.*, HMGN1 at the amino terminus and HMGB1 box A at the carboxyl terminus), and McaCas9-HN1H1G fusion (*i.e.*, HMGN1 at the amino terminus and histone H1 central globular motif at the carboxyl terminus). The target site of each locus is presented in Table 6. Error bars show mean ± SD (n = 3 biological replicates).

[0012] FIG. **2B** presents the cleavage efficiency (as the percent of indels) of PexCas9, PexCas9-HN1HB1 fusion (*i.e.*, HMGN1 at the amino terminus and HMGB1 box A at the carboxyl terminus), and PexCas9-HN1H1G fusion (*i.e.*, HMGN1 at the amino terminus and histone H1 central globular motif at the carboxyl terminus). The target site of each locus is presented in Table 6. Error bars show mean ± SD (n = 3 biological replicates).

[0013] FIG. **2C** presents the cleavage efficiency (as the percent of indels) of BsmCas9, BsmCas9-HN1HB1 fusion (*i.e.*, HMGN1 at the amino terminus and HMGB1 box A at the carboxyl terminus), and BsmCas9-HN1H1G fusion (*i.e.*, HMGN1 at the amino terminus and histone H1 central globular motif at the carboxyl terminus). The target site of each locus is presented in Table 6. Error bars show mean ± SD (n = 3 biological replicates).

[0014] FIG. **2D** presents the cleavage efficiency (as the percent of indels) of LrhCas9, LrhCas9-HN1HB1 fusion (*i.e.*, HMGN1 at the amino terminus and HMGB1 box A at the carboxyl terminus), and LrhCas9-HN1H1G fusion (*i.e.*, HMGN1 at the amino terminus and histone H1 central globular motif at the carboxyl terminus). The target site of each locus is presented in Table 6. Error bars show mean ± SD (n = 3 biological replicates).

[0015] FIG. **3** shows off-target activities (as the percent of indels) of control Cas9 and Cas9-CMM fusion nucleases. Error bars show mean ± SD (n = 3 biological replicates).

DETAILED DESCRIPTION

[0016] The present disclosure provides orthologous Cas9 systems that use alternate PAMs for target DNA binding, thereby increasing genome coverage density. For example, some of these alternate PAMs comprise A and/or T residues, and other alternate PAMS are GC-rich. As such, the engineered Cas9 systems that utilize these alternate PAMs enable targeted genome editing or genome modification of previously inaccessible genomic loci.

(I) Engineered Cas9 Systems

[0017] One aspect of the present disclosure provides engineered Cas9 systems comprising engineered Cas9 proteins and engineered guide RNAs, wherein each engineered guide RNA is designed to complex with a specific engineered Cas9 protein. Each engineered guide RNA comprises a 5' guide sequence designed to hybridize with a target sequence in a double-stranded sequence, wherein the target sequence is 5' to a protospacer adjacent motif (PAM) and the PAM has a sequence as listed in Table A. These engineered Cas9 systems do not occur naturally.

(a) Engineered Cas9 Proteins

[0018] The engineered Cas9 protein comprises at least one amino acid substitution, insertion, or deletion relative to its wild-type counterpart. Cas9 protein is the single effector protein in type II CRISPR systems, which are present in various bacteria. The engineered Cas9 protein disclosed herein can be from *Acaryochloris sp., Acetohalobium sp., Acidaminococcus sp., Acidithiobacillus sp., Acidothermus* sp., *Akkermansia sp., Alicyclobacillus sp., Allochromatium sp., Ammonifex sp., Anabaena sp., Arthrospira sp., Bacillus sp., Bifidobacterium sp., Burkholderiales sp., Caldicelulosiruptor sp., Campylobacter sp., Candidatus sp., Clostridium sp.,*

Corynebacterium sp., Crocosphaera sp., Cyanothece sp., Exiguobacterium sp., Finegoldia sp., Francisella sp., *Ktedonobacter sp., Lachnospiraceae sp., Lactobacillus sp., Lyngbya sp., Marinobacter sp., Methanohalobium sp., Microscilla sp., Microcoleus sp., Microcystis sp., Mycoplasma sp., Natranaerobius sp., Neisseria sp., Nitratifractor sp., Nitrosococcus sp., Nocardiopsis sp., Nodularia sp., Nostoc sp., Oenococcus sp., Oscillatoria sp., Parasutterella sp., Pelotomaculum sp., Petrotoga sp., Polaromonas sp., Prevotella sp., Pseudoalteromonas sp., Ralstonia sp., Staphylococcus sp., Streptococcus sp., Streptomyces sp., Streptosporangium sp., Synechococcus sp., Thermosipho sp.*, *Verrucomicrobia s*p., and *Wolinella* sp.

[0019] In certain embodiments, the engineered Cas9 protein disclosed herein is from *Acidothermus* sp., *Akkermansia sp., Alicyclobacillus sp., Bacillus sp., Bifidobacterium sp., Burkholderiales sp., Corynebacterium sp., Lactobacillus sp., Mycoplasma sp., Nitratifractor sp., Oenococcus sp., Parasutterella sp., Ralstonia sp., or Wolinella* sp.

[0020] In specific embodiments, the engineered Cas9 protein disclosed herein is from *Acidothermus cellulolyticus* (Ace), *Akkermansia glycaniphila* (Agl), *Akkermansia muciniphila* (Amu), *Alicyclobacillus hesperidum (*Ahe), *Bacillus smithii (*Bsm), *Bifidobacterium bombi* (Bbo), *Corynebacterium diphtheria* (Cdi), *Lactobacillus rhamnosus* (Lrh), *Mycoplasma canis* (Mca), *Mycoplasma gallisepticum* (Mga), *Nitratifractor salsuginis* (Nsa), *Oenococcus kitaharae (*Oki), *Parasutterella excrementihominis* (Pex), *Ralstonia syzygii* (Rsy), or *Wolinella succinogenes* (Wsu).

[0021] Wild-type Cas9 proteins comprise two nuclease domains, *i.e.*, RuvC and HNH domains, each of which cleaves one strand of a double-stranded sequence. Cas9 proteins also comprise REC domains that interact with the guide RNA (*e.g.*, REC1, REC2) or the RNA/DNA heteroduplex (*e.g.*, REC3), and a domain that interacts with the protospacer-adjacent motif (PAM) (*i.e.*, PAM-interacting domain).

[0022] The Cas9 protein can be engineered to comprise one or more modifications (*i.e.*, a substitution of at least one amino acid, a deletion of at least one amino acid, an insertion of at least one amino acid) such that the Cas9 protein has altered activity, specificity, and/or stability.

[0023] For example, Cas9 protein can be engineered by one or more mutations and/or deletions to inactivate one or both of the nuclease domains. Inactivation of one nuclease domain generates a Cas9 protein that cleaves one strand of a double-stranded sequence (*i.e.*, a Cas9 nickase). The RuvC domain can be inactivated by mutations such as D10A, D8A, E762A, and/or D986A, and the HNH domain can be inactivated by mutations such as H840A, H559A, N854A, N856A, and/or N863A (with reference to the numbering system of *Streptococcus pyogenes* Cas9, SpyCas9). Inactivation of both nuclease domains generates a Cas9 protein having no cleavage activity (*i.e.*, a catalytically inactive or dead Cas9).

[0024] The Cas9 protein can also be engineered by one or more amino acid substitutions, deletions, and/or insertions to have improved targeting specificity, improved fidelity, altered PAM specificity, decreased off-target effects, and/or increased stability. Non-limiting examples of one or more mutations that improve targeting specificity, improve fidelity, and/or decrease off-target effects include N497A, R661A, Q695A, K810A, K848A, K855A, Q926A, K1003A, R1060A, and/or D1135E (with reference to the numbering system of SpyCas9).

(i) Heterologous domains

[0025] The Cas9 protein can be engineered to comprise at least one heterologous domain, *i.e.*, Cas9 is fused to one or more heterologous domains. In situations in which two or more heterologous domains are fused with Cas9, the two or more heterologous domains can be the same or they can be different. The one or more heterologous domains can be fused to the N terminal end, the C terminal end, an internal location, or combination thereof. The fusion can be direct via a chemical bond, or the linkage can be indirect via one or more linkers. In various embodiments, the heterologous domain can be a nuclear localization signal, a cell-penetrating domain, a marker domain, a chromatin disrupting domain, an epigenetic modification domain (*e.g.*, a cytidine deaminase domain, a histone acetyltransferase domain, and the like), a transcriptional regulation domain, an RNA aptamer binding domain, or a non-Cas9 nuclease domain.

[0026] In some embodiments the one or more heterologous domains can be a nuclear localization signal (NLS). Non-limiting examples of nuclear localization signals include PKKKRKV (SEQ ID NO:78), PKKKRRV (SEQ ID NO:79), KRPAATKKAGQAKKKK (SEQ ID NO:80), YGRKKRRQRRR (SEQ ID NO:81), RKKRRQRRR (SEQ ID NO:82), PAAKRVKLD (SEQ ID NO:83), RQRRNELKRSP (SEQ ID NO:84), VSRKRPRP (SEQ ID NO:85), PPKKARED (SEQ ID NO:86), PQPKKKPL (SEQ ID NO:87), SALIKKKKKMAP (SEQ ID NO:88), PKQKKRK (SEQ ID NO:89), RKLKKKIKKL (SEQ ID NO:90), REKKKFLKRR (SEQ ID NO:91), KRKGDEVDGVDEVAKKKSKK (SEQ ID NO:92), RKCLQAGMNLEARKTKK (SEQ ID NO:93), NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (SEQ ID NO:94), and RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV (SEQ ID NO:95).

[0027] In other embodiments, the one or more heterologous domains can be a cell-penetrating domain. Examples of suitable cell-penetrating domains include, without limit, GRKKRRQRRRPPQPKKKRKV (SEQ ID NO:96), PLSSIFSRIGDPPKKKRKV (SEQ ID NO:97), GALFLGWLGAAGSTMGAPKKKRKV (SEQ ID NO:98), GALFLGFLGAAGSTMGAWSQPKKKRKV (SEQ ID NO:99), KETWWETWWTEWSQPKKKRKV (SEQ ID NO:100), YARAAARQARA (SEQ ID NO:101), THRLPRRRRRR (SEQ ID NO:102), GGRRARRRRRR (SEQ ID NO:103), RRQRRTSKLMKR (SEQ ID NO:104), GWTLNSAGYLLGKINLKALAALAKKIL (SEQ ID NO:105), KALAWEAKLAKALAKALAKHLAKALAKALKCEA (SEQ ID NO:106), and RQIKIWFQNRRMKWKK (SEQ ID NO:107).

[0028] In alternate embodiments, the one or more heterologous domains can be a marker domain. Marker domains include fluorescent proteins and purification or epitope tags. Suitable fluorescent proteins include, without limit, green fluorescent proteins (*e.g.*, GFP, eGFP, GFP-2, tagGFP, turboGFP, Emerald, Azami Green, Monomeric Azami Green, CopGFP, AceGFP, ZsGreen1), yellow fluorescent proteins (*e.g.*, YFP, EYFP, Citrine, Venus, YPet, PhiYFP, ZsYellow1), blue fluorescent proteins (*e.g.*, BFP, EBFP, EBFP2, Azurite, mKalama1, GFPuv, Sapphire, T-sapphire), cyan fluorescent proteins (*e.g.,* ECFP, Cerulean, CyPet, AmCyan1, Midoriishi-Cyan), red fluorescent proteins (*e.g.*, mKate, mKate2, mPlum, DsRed monomer, mCherry, mRFP1,

DsRed-Express, DsRed2, DsRed-Monomer, HcRed-Tandem, HcRed1, AsRed2, eqFP611, mRasberry, mStrawberry, Jred), orange fluorescent proteins (*e.g.*, mOrange, mKO, Kusabira-Orange, Monomeric Kusabira-Orange, mTangerine, tdTomato), or combinations thereof. The marker domain can comprise tandem repeats of one or more fluorescent proteins (*e.g.*, Suntag). Non-limiting examples of suitable purification or epitope tags include 6xHis (SEQ ID NO: 134), FLAG®, HA, GST, Myc, SAM, and the like. Non-limiting examples of heterologous fusions which facilitate detection or enrichment of CRISPR complexes include streptavidin (Kipriyanov et al., Human Antibodies, 1995, 6(3):93-101.), avidin (Airenne et al., Biomolecular Engineering, 1999, 16(1-4):87-92), monomeric forms of avidin (Laitinen et al., Journal of Biological Chemistry, 2003, 278(6):4010-4014), peptide tags which facilitate biotinylation during recombinant production (Cull et al., Methods in Enzymology, 2000, 326:430-440).

[0029] In still other embodiments, the one or more heterologous domain can be a chromatin modulating motif (CMM). Non-limiting examples of CMMs include nucleosome interacting peptides derived from high mobility group (HMG) proteins (*e.g.*, HMGB1, HMGB2, HMGB3, HMGN1, HMGN2, HMGN3a, HMGN3b, HMGN4, and HMGN5 proteins), the central globular domain of histone H1 variants (*e.g.*, histone H1.0, H1.1, H1.2, H1.3, H1.4, H1.5, H1.6, H1.7, H1.8, H1.9, and H.1.10), or DNA binding domains of chromatin remodeling complexes (*e.g.*, SWI/SNF (SWItch/Sucrose Non-Fermentable), ISWI (Imitation SWItch), CHD (Chromodomain-Helicase-DNA binding), Mi-2/NuRD (Nucleosome Remodeling and Deacetylase), INO80, SWR1, and RSC complexes. In other embodiments, CMMs also can be derived from topoisomerases, helicases, or viral proteins. The source of the CMM can and will vary. CMMs can be from humans, animals (*i.e.*, vertebrates and invertebrates), plants, algae, or yeast. Non-limiting examples of specific CMMs are listed in the table below. Persons of skill in the art can readily identify homologs in other species and/or the relevant fusion motif therein.

Protein	Accession No.	Fusion Motif
Human HMGN1	P05114	Full length

Human HMGN2	P05204	Full length
Human HMGN3a	Q15651	Full length
Human HMGN3b	Q15651-2	Full length
Human HMGN4	O00479	Full length
Human HMGN5	P82970	Nucleosome binding motif
Human HMGB1	P09429	Box A
Human histone H1.0	P07305	Globular motif
Human histone H1.2	P16403	Globular motif
Human CHD1	O14646	DNA binding motif
Yeast CHD1	P32657	DNA binding motif
Yeast ISWI	P38144	DNA binding motif
Human TOP1	P11387	DNA binding motif
Human herpesvirus 8 LANA	J9QSF0	Nucleosome binding motif
Human CMV IE1	P13202	Chromatin tethering motif
M. leprae DNA helicase	P40832	HhH binding motif

[0030] In yet other embodiments, the one or more heterologous domains can be an epigenetic modification domain. Non-limiting examples of suitable epigenetic modification domains include those with DNA deamination (*e.g.*, cytidine deaminase, adenosine deaminase, guanine deaminase), DNA methyltransferase activity (*e.g.*, cytosine methyltransferase), DNA demethylase activity, DNA amination, DNA oxidation activity, DNA helicase activity, histone acetyltransferase (HAT) activity (*e.g.*, HAT domain derived from E1A binding protein p300), histone deacetylase activity, histone methyltransferase activity, histone demethylase activity, histone kinase activity, histone phosphatase activity, histone ubiquitin ligase activity, histone deubiquitinating activity, histone adenylation activity, histone deadenylation activity, histone SUMOylating activity, histone deSUMOylating activity, histone ribosylation activity, histone deribosylation activity, histone myristoylation activity, histone demyristoylation activity, histone citrullination activity, histone alkylation activity, histone dealkylation activity, or histone oxidation activity. In specific embodiments, the epigenetic modification domain

can comprise cytidine deaminase activity, adenosine deaminase activity, histone acetyltransferase activity, or DNA methyltransferase activity.

[0031] In other embodiments, the one or more heterologous domains can be a transcriptional regulation domain (*i.e.*, a transcriptional activation domain or transcriptional repressor domain). Suitable transcriptional activation domains include, without limit, herpes simplex virus VP16 domain, VP64 (*i.e.*, four tandem copies of VP16), VP160 (*i.e.*, ten tandem copies of VP16), NFκB p65 activation domain (p65) , Epstein-Barr virus R transactivator (Rta) domain, VPR (*i.e.*, VP64+p65+Rta), p300-dependent transcriptional activation domains, p53 activation domains 1 and 2, heat-shock factor 1 (HSF1) activation domains, Smad4 activation domains (SAD), cAMP response element binding protein (CREB) activation domains, E2A activation domains, nuclear factor of activated T-cells (NFAT) activation domains, or combinations thereof. Non-limiting examples of suitable transcriptional repressor domains include Kruppel-associated box (KRAB) repressor domains, Mxi repressor domains, inducible cAMP early repressor (ICER) domains, YY1 glycine rich repressor domains, Sp1-like repressors, E(spl) repressors, IκB repressors, Sin3 repressors, methyl-CpG binding protein 2 (MeCP2) repressors, or combinations thereof. Transcriptional activation or transcriptional repressor domains can be genetically fused to the Cas9 protein or bound via noncovalent protein-protein, protein-RNA, or protein-DNA interactions.

[0032] In further embodiments, the one or more heterologous domains can be an RNA aptamer binding domain (Konermann *et al.*, Nature, 2015, 517(7536):583-588; Zalatan *et al.*, Cell, 2015, 160(1-2):339-50). Examples of suitable RNA aptamer protein domains include MS2 coat protein (MCP), PP7 bacteriophage coat protein (PCP), Mu bacteriophage Com protein, lambda bacteriophage N22 protein, stem-loop binding protein (SLBP), Fragile X mental retardation syndrome-related protein 1 (FXR1), proteins derived from bacteriophage such as AP205, BZ13, f1, f2, fd, fr, ID2, JP34/GA, JP501, JP34, JP500, KU1, M11, M12, MX1, NL95, PP7, ϕCb5, ϕCb8r, ϕCb12r, ϕCb23r, Qβ, R17, SP-β, TW18, TW19, and VK, fragments thereof, or derivatives thereof.

[0033] In yet other embodiments, the one or more heterologous domains can be a non-Cas9 nuclease domain. Suitable nuclease domains can be obtained from

any endonuclease or exonuclease. Non-limiting examples of endonucleases from which a nuclease domain can be derived include, but are not limited to, restriction endonucleases and homing endonucleases. In some embodiments, the nuclease domain can be derived from a type II-S restriction endonuclease. Type II-S endonucleases cleave DNA at sites that are typically several base pairs away from the recognition/binding site and, as such, have separable binding and cleavage domains. These enzymes generally are monomers that transiently associate to form dimers to cleave each strand of DNA at staggered locations. Non-limiting examples of suitable type II-S endonucleases include BfiI, BpmI, BsaI, BsgI, BsmBI, BsmI, BspMI, FokI, MboII, and SapI. In some embodiments, the nuclease domain can be a FokI nuclease domain or a derivative thereof. The type II-S nuclease domain can be modified to facilitate dimerization of two different nuclease domains. For example, the cleavage domain of FokI can be modified by mutating certain amino acid residues. By way of non-limiting example, amino acid residues at positions 446, 447, 479, 483, 484, 486, 487, 490, 491, 496, 498, 499, 500, 531, 534, 537, and 538 of FokI nuclease domains are targets for modification. In specific embodiments, the FokI nuclease domain can comprise a first FokI half- domain comprising Q486E, I499L, and/or N496D mutations, and a second FokI half-domain comprising E490K, I538K, and/or H537R mutations.

[0034] The one or more heterologous domains can be linked directly to the Cas9 protein via one or more chemical bonds (*e.g.*, covalent bonds), or the one or more heterologous domains can be linked indirectly to the Cas9 protein via one or more linkers.

[0035] A linker is a chemical group that connects one or more other chemical groups via at least one covalent bond. Suitable linkers include amino acids, peptides, nucleotides, nucleic acids, organic linker molecules (*e.g.*, maleimide derivatives, N-ethoxybenzylimidazole, biphenyl-3,4′,5-tricarboxylic acid, p-aminobenzyloxycarbonyl, and the like), disulfide linkers, and polymer linkers (*e.g.*, PEG). The linker can include one or more spacing groups including, but not limited to alkylene, alkenylene, alkynylene, alkyl, alkenyl, alkynyl, alkoxy, aryl, heteroaryl, aralkyl, aralkenyl, aralkynyl and the like. The linker can be neutral, or carry a positive or

negative charge. Additionally, the linker can be cleavable such that the linker's covalent bond that connects the linker to another chemical group can be broken or cleaved under certain conditions, including pH, temperature, salt concentration, light, a catalyst, or an enzyme. In some embodiments, the linker can be a peptide linker. The peptide linker can be a flexible amino acid linker (*e.g.*, comprising small, non-polar or polar amino acids). Non-limiting examples of flexible linkers include LEGGGS (SEQ ID NO:108), TGSG (SEQ ID NO:109), GGSGGGSG (SEQ ID NO:110), $(GGGGS)_{1-4}$ (SEQ ID NO:111), and $(Gly)_{6-8}$ (SEQ ID NO:112). Alternatively, the peptide linker can be a rigid amino acid linker. Such linkers include $(EAAAK)_{1-4}$ (SEQ ID NO:113), $A(EAAAK)_{2-5}A$ (SEQ ID NO:114), PAPAP (SEQ ID NO:115), and $(AP)_{6-8}$ (SEQ ID NO:116). Additional examples of suitable linkers are well known in the art and programs to design linkers are readily available (Crasto *et al.*, Protein Eng., 2000, 13(5):309-312).

[0036] In some embodiments, the engineered Cas9 proteins can be produced recombinantly in cell-free systems, bacterial cells, or eukaryotic cells and purified using standard purification means. In other embodiments, the engineered Cas9 proteins are produced *in vivo* in eukaryotic cells of interest from nucleic acids encoding the engineered Cas9 proteins (see section (II) below).

[0037] In embodiments in which the engineered Cas9 protein comprises nuclease or nickase activity, the engineered Cas9 protein can further comprise at least one nuclear localization signal, cell-penetrating domain, and/or marker domain, as well as at least one chromatin disrupting domain. In embodiments in which the engineered Cas9 protein is linked to an epigenetic modification domain, the engineered Cas9 protein can further comprise at least one nuclear localization signal, cell-penetrating domain, and/or marker domain, as well as at least one chromatin disrupting domain. Furthermore, in embodiments in which the engineered Cas9 protein is linked to a transcriptional regulation domain, the engineered Cas9 protein can further comprise at least one nuclear localization signal, cell-penetrating domain, and/or marker domain, as well as at least one chromatin disrupting domain and/or at least one RNA aptamer binding domain.

(ii) Specific engineered Cas9 proteins

[0038] In specific embodiments, the engineered Cas9 protein is from *Bacillus smithii*, *Lactobacillus rhamnosus*, *Parasutterella excrementihominis*, *Mycoplasma canis*, *Mycoplasma gallisepticum*, *Akkermansia glycaniphila*, *Akkermansia muciniphila*, *Oenococcus kitaharae*, *Bifidobacterium bombi*, *Acidothermus cellulolyticus*, *Alicyclobacillus hesperidum*, *Wolinella succinogenes*, *Nitratifractor salsuginis*, *Ralstonia syzygii*, or *Corynebacterium diphtheria* and is linked to at least one NLS. In some iterations, the engineered Cas9 protein can have at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99% sequence identity to SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, or 30. In certain embodiments, the engineered Cas9 protein can have at least about 95% sequence identity to SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, or 30. In other iterations, the engineered Cas9 protein has the amino acid sequence of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, or 30.

[0039] In other embodiments, the engineered Cas9 protein can be a *Bacillus smithii*, *Lactobacillus rhamnosus*, *Parasutterella excrementihominis*, *Mycoplasma canis*, *Mycoplasma gallisepticum*, *Akkermansia glycaniphila*, *Akkermansia muciniphila*, *Oenococcus kitaharae*, *Bifidobacterium bombi*, *Acidothermus cellulolyticus*, *Alicyclobacillus hesperidum*, *Wolinella succinogenes*, *Nitratifractor salsuginis*, *Ralstonia syzygii*, or *Corynebacterium diphtheria* Cas9 protein linked to at least one chromatin modulating motif (CMM). The linkage between the Cas9 protein and the CMM can be direct or via a linker. The Cas9-CMM fusion protein can further comprise at least one NLS. In particular embodiments, the Cas9-CMM fusion protein can have at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99% sequence identity to SEQ ID NO:117, 118, 119, 1200, 121, 122, 123, or 124. In certain embodiments, the Cas9-CMM fusion protein can have at least about 95% sequence identity to SEQ ID NO:117, 118, 119, 120, 121, 122, 123, or 124. In specific iterations, the Cas9-CMM fusion protein has the amino acid sequence of SEQ ID NO:117, 118, 119, 120, 121, 122, 123, or 124.

(b) Engineered guide RNAs

[0040] The engineered guide RNAs is designed to complex with a specific engineered Cas9 protein. A guide RNA comprises (i) a CRISPR RNA (crRNA) that contains a guide sequence at the 5' end that hybridizes with a target sequence and (ii) a transacting crRNA (tracrRNA) sequence that recruits the Cas9 protein. The crRNA guide sequence of each guide RNA is different (*i.e.*, is sequence specific). The tracrRNA sequence is generally the same in guide RNAs designed to complex with a Cas9 protein from a particular bacterial species.

[0041] The crRNA guide sequence is designed to hybridize with a target sequence (*i.e.*, protospacer) in a double-stranded sequence. In general, the complementarity between the crRNA and the target sequence is at least 80%, at least 85%, at least 90%, at least 95%, or at least 99%. In specific embodiments, the complementarity is complete (*i.e.*, 100%). In various embodiments, the length of the crRNA guide sequence can range from about 15 nucleotides to about 25 nucleotides. For example, the crRNA guide sequence can be about 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 nucleotides in length. In specific embodiments, the crRNA is about 19, 20, or 21 nucleotides in length. In one embodiment, the crRNA guide sequence has a length of 20 nucleotides.

[0042] The guide RNA comprises repeat sequence that forms at least one stem loop structure, which interacts with the Cas9 protein, and 3' sequence that remains single-stranded. The length of each loop and stem can vary. For example, the loop can range from about 3 to about 10 nucleotides in length, and the stem can range from about 6 to about 20 base pairs in length. The stem can comprise one or more bulges of 1 to about 10 nucleotides. The length of the single-stranded 3' region can vary. The tracrRNA sequence in the engineered guide RNA generally is based upon the coding sequence of wild type tracrRNA in the bacterial species of interest. The wild-type sequence can be modified to facilitate secondary structure formation, increased secondary structure stability, facilitate expression in eukaryotic cells, and so forth. For example, one or more nucleotide changes can be introduced into the guide RNA coding sequence (see Example 3, below). The tracrRNA sequence can range in length from

about 50 nucleotides to about 300 nucleotides. In various embodiments, the tracrRNA can range in length from about 50 to about 90 nucleotides, from about 90 to about 110 nucleotides, from about 110 to about 130 nucleotides, from about 130 to about 150 nucleotides, from about 150 to about 170 nucleotides, from about 170 to about 200 nucleotides, from about 200 to about 250 nucleotides, or from about 250 to about 300 nucleotides.

[0043] In general, the engineered guide RNA is a single molecule (*i.e.*, a single guide RNA or sgRNA), wherein the crRNA sequence is linked to the tracrRNA sequence. In some embodiments, however, the engineered guide RNA can be two separate molecules. A first molecule comprising the crRNA that contains 3' sequence (comprising from about 6 to about 20 nucleotides) that is capable of base pairing with the 5' end of a second molecule, wherein the second molecule comprises the tracrRNA that contains 5' sequence (comprising from about 6 to about 20 nucleotides) that is capable of base pairing with the 3' end of the first molecule.

[0044] In some embodiments, the tracrRNA sequence of the engineered guide RNA can be modified to comprise one or more aptamer sequences (Konermann *et al.*, Nature, 2015, 517(7536):583-588; Zalatan *et al.*, Cell, 2015, 160(1-2):339-50). Suitable aptamer sequences include those that bind adaptor proteins chosen from MCP, PCP, Com, SLBP, FXR1, AP205, BZ13, f1, f2, fd, fr, ID2, JP34/GA, JP501, JP34, JP500, KU1, M11, M12, MX1, NL95, PP7, ϕCb5, ϕCb8r, ϕCb12r, ϕCb23r, Qβ, R17, SP-β, TW18, TW19, VK, fragments thereof, or derivatives thereof. Those of skill in the art appreciate that the length of the aptamer sequence can vary.

[0045] In other embodiments, the guide RNA can further comprise at least one detectable label. The detectable label can be a fluorophore (*e.g.*, FAM, TMR, Cy3, Cy5, Texas Red, Oregon Green, Alexa Fluors, Halo tags, or suitable fluorescent dye), a detection tag (*e.g.*, biotin, digoxigenin, and the like), quantum dots, or gold particles.

[0046] The guide RNA can comprise standard ribonucleotides and/or modified ribonucleotides. In some embodiment, the guide RNA can comprise standard or modified deoxyribonucleotides. In embodiments in which the guide RNA is enzymatically synthesized (*i.e.*, *in vivo* or *in vitro*), the guide RNA generally comprises

standard ribonucleotides. In embodiments in which the guide RNA is chemically synthesized, the guide RNA can comprise standard or modified ribonucleotides and/or deoxyribonucleotides. Modified ribonucleotides and/or deoxyribonucleotides include base modifications (*e.g.*, pseudouridine, 2-thiouridine, N6-methyladenosine, and the like) and/or sugar modifications (*e.g.*, 2'-O-methy, 2'-fluoro, 2'-amino, locked nucleic acid (LNA), and so forth). The backbone of the guide RNA can also be modified to comprise phosphorothioate linkages, boranophosphate linkages, or peptide nucleic acids.

[0047] In specific embodiments, the engineered guide RNA has at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99% sequence identity to SEQ ID NO:31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, or 45. In some embodiments, the engineered Cas9 guide RNA has the sequence of SEQ ID NO:31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, or 45.

(c) PAM Sequence

[0048] The engineered Cas9 systems detailed above target specific sequences in double-stranded DNA that are located upstream of novel PAM sequences. The PAM sequences preferred by the engineered Cas9 systems were identified *in vitro* using a library of degenerate PAMS (see Example 1 and FIG. **1**), and confirmed by sequencing after genome editing experiments (see Example 2). The PAM for each of the engineered Cas9 system disclosed herein is presented in Table A, below.

Table A. PAM Sequences	
Engineered Cas9system	PAM (5'-3')*
Bacillus smithii Cas9 (BsmCas9)	NNNNCAAA
Lactobacillus rhamnosus Cas9 (LrhCas9)	NGAAA
Parasutterella excrementihominis Cas9 (PexCas9)	NGG

Mycoplasma canis Cas9 (McaCas9)	NNGG
Mycoplasma gallisepticum Cas9 (MgaCas9)	NNAAT
Akkermansia glycaniphila Cas9 (AglCas9)	NNNRTA
Akkermansia muciniphila Cas9 (AmuCas9)	MMACCA
Oenococcus kitaharae Cas9 (OkiCas9)	NNG
Bifidobacterium bombi Cas9 (BboCas9)	NNNNGRY
Acidothermus cellulolyticus Cas9 (AceCas9)	NGG
Alicyclobacillus hesperidum Cas9 (AheCas9)	NGG
Wolinella succinogenes Cas9 (WsuCas9)	NGG
Nitratifractor salsuginis Cas9 (NsaCas9)	NRGNK
Ralstonia syzygii Cas9 (RsyCas9)	GGGRG
Corynebacterium diphtheria Cas9 (CdiCas9)	NNAMMMC

* K is G or T; M is A or C; R is A or G; Y is C or T; and N is A, C, G, or T.

(II) Nucleic Acids

[0049] A further aspect of the present disclosure provides nucleic acids encoding the engineered Cas9 systems described above in section (I). The systems can be encoded by single nucleic acids or multiple nucleic acids. The nucleic acids can be DNA or RNA, linear or circular, single-stranded or double-stranded. The RNA or DNA can be codon optimized for efficient translation into protein in the eukaryotic cell of interest. Codon optimization programs are available as freeware or from commercial sources.

[0050] In some embodiments, nucleic acid encodes a protein having at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99% sequence identity to the amino acid sequence of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, or 30. In certain embodiments, the nucleic acid encoding the engineered Cas9 protein can have at least about 75%, at

least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99% sequence identity to the DNA sequence of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, or 29. In certain embodiments, the DNA encoding the engineered Cas9 protein has the DNA sequence of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, or 29. In additional embodiments, the nucleic acid encodes a protein having at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99% sequence identity to the amino acid sequence of SEQ ID NO:117, 118, 119, 120, 121, 122, 123, or 124.

[0051] In some embodiments, the nucleic acid encoding the engineered Cas9 protein can be RNA. The RNA can be enzymatically synthesized *in vitro*. For this, DNA encoding the engineered Cas9 protein can be operably linked to a promoter sequence that is recognized by a phage RNA polymerase for *in vitro* RNA synthesis. For example, the promoter sequence can be a T7, T3, or SP6 promoter sequence or a variation of a T7, T3, or SP6 promoter sequence. The DNA encoding the engineered protein can be part of a vector, as detailed below. In such embodiments, the *in vitro*-transcribed RNA can be purified, capped, and/or polyadenylated. In other embodiments, the RNA encoding the engineered Cas9 protein can be part of a self-replicating RNA (Yoshioka *et al.*, Cell Stem Cell, 2013, 13:246-254). The self-replicating RNA can be derived from a noninfectious, self-replicating Venezuelan equine encephalitis (VEE) virus RNA replicon, which is a positive-sense, single-stranded RNA that is capable of self-replicating for a limited number of cell divisions, and which can be modified to code proteins of interest (Yoshioka *et al.*, Cell Stem Cell, 2013, 13:246-254).

[0052] In other embodiments, the nucleic acid encoding the engineered Cas9 protein can be DNA. The DNA coding sequence can be operably linked to at least one promoter control sequence for expression in the cell of interest. In certain embodiments, the DNA coding sequence can be operably linked to a promoter sequence for expression of the engineered Cas9 protein in bacterial (*e.g.*, *E. coli*) cells or eukaryotic (*e.g.*, yeast, insect, or mammalian) cells. Suitable bacterial promoters include, without limit, T7 promoters, *lac* operon promoters, *trp* promoters, *tac* promoters (which are hybrids of *trp* and *lac* promoters), variations of any of the foregoing, and

combinations of any of the foregoing. Non-limiting examples of suitable eukaryotic promoters include constitutive, regulated, or cell- or tissue-specific promoters. Suitable eukaryotic constitutive promoter control sequences include, but are not limited to, cytomegalovirus immediate early promoter (CMV), simian virus (SV40) promoter, adenovirus major late promoter, Rous sarcoma virus (RSV) promoter, mouse mammary tumor virus (MMTV) promoter, phosphoglycerate kinase (PGK) promoter, elongation factor (ED1)-alpha promoter, ubiquitin promoters, actin promoters, tubulin promoters, immunoglobulin promoters, fragments thereof, or combinations of any of the foregoing. Examples of suitable eukaryotic regulated promoter control sequences include without limit those regulated by heat shock, metals, steroids, antibiotics, or alcohol. Non-limiting examples of tissue-specific promoters include B29 promoter, CD14 promoter, CD43 promoter, CD45 promoter, CD68 promoter, desmin promoter, elastase-1 promoter, endoglin promoter, fibronectin promoter, Flt-1 promoter, GFAP promoter, GPIIb promoter, ICAM-2 promoter, INF-β promoter, Mb promoter, NphsI promoter, OG-2 promoter, SP-B promoter, SYN1 promoter, and WASP promoter. The promoter sequence can be wild type or it can be modified for more efficient or efficacious expression. In some embodiments, the DNA coding sequence also can be linked to a polyadenylation signal (*e.g.*, SV40 polyA signal, bovine growth hormone (BGH) polyA signal, *etc.*) and/or at least one transcriptional termination sequence. In some situations, the engineered Cas9 protein can be purified from the bacterial or eukaryotic cells.

[0053] In still other embodiments, the engineered guide RNA can be encoded by DNA. In some instances, the DNA encoding the engineered guide RNA can be operably linked to a promoter sequence that is recognized by a phage RNA polymerase for *in vitro* RNA synthesis. For example, the promoter sequence can be a T7, T3, or SP6 promoter sequence or a variation of a T7, T3, or SP6 promoter sequence. In other instances, the DNA encoding the engineered guide RNA can be operably linked to a promoter sequence that is recognized by RNA polymerase III (Pol III) for expression in eukaryotic cells of interest. Examples of suitable Pol III promoters include, but are not limited to, mammalian U6, U3, H1, and 7SL RNA promoters.

[0054] In various embodiments, the nucleic acid encoding the engineered Cas9 protein can be present in a vector. In some embodiments, the vector can further comprise nucleic acid encoding the engineered guide RNA. Suitable vectors include plasmid vectors, viral vectors, and self-replicating RNA (Yoshioka *et al.*, Cell Stem Cell, 2013, 13:246-254). In some embodiments, the nucleic acid encoding the complex or fusion protein can be present in a plasmid vector. Non-limiting examples of suitable plasmid vectors include pUC, pBR322, pET, pBluescript, and variants thereof. In other embodiments, the nucleic acid encoding the complex or fusion protein can be part of a viral vector (*e.g.*, lentiviral vectors, adeno-associated viral vectors, adenoviral vectors, and so forth). The plasmid or viral vector can comprise additional expression control sequences (*e.g.*, enhancer sequences, Kozak sequences, polyadenylation sequences, transcriptional termination sequences, *etc.*), selectable marker sequences (*e.g.*, antibiotic resistance genes), origins of replication, and the like. Additional information about vectors and use thereof can be found in "Current Protocols in Molecular Biology" Ausubel *et al.*, John Wiley & Sons, New York, 2003 or "Molecular Cloning: A Laboratory Manual" Sambrook & Russell, Cold Spring Harbor Press, Cold Spring Harbor, NY, 3rd edition, 2001.

(III) Eukaryotic Cells

[0055] Another aspect of the present disclosure comprises eukaryotic cells comprising at least one engineered Cas9 system as detailed above in section (I) and/or at least one nucleic acid encoding and engineered Cas9 protein and/or engineered guide RNA as detailed above in section (II).

[0056] The eukaryotic cell can be a human cell, a non-human mammalian cell, a non-mammalian vertebrate cell, an invertebrate cell, a plant cell, or a single cell eukaryotic organism. Examples of suitable eukaryotic cells are detailed below in section (IV)(c). The eukaryotic cell can be *in vitro*, *ex vivo*, or *in vivo*.

(IV) Methods for Modifying Chromosomal Sequences

[0057] A further aspect of the present disclosure encompasses methods for modifying a chromosomal sequence in eukaryotic cells. In general, the methods comprise introducing into the eukaryotic cell of interest at least one engineered Cas9 system as detailed above in section (I) and/or at least one nucleic acid encoding said engineered Cas9 system as detailed above in section (II).

[0058] In embodiments in which the engineered Cas9 protein comprises nuclease or nickase activity, the chromosomal sequence modification can comprise a substitution of at least one nucleotide, a deletion of at least one nucleotide, an insertion of at least one nucleotide. In some iterations, the method comprises introducing into the eukaryotic cell one engineered Cas9 system comprising nuclease activity or two engineered Cas9 systems comprising nickase activity and no donor polynucleotide, such that the engineered Cas9 system or systems introduce a double-stranded break in the target site in the chromosomal sequence and repair of the double-stranded break by cellular DNA repair processes introduces at least one nucleotide change (*i.e.*, indel), thereby inactivating the chromosomal sequence (*i.e.*, gene knock-out). In other iterations, the method comprises introducing into the eukaryotic cell one engineered Cas9 system comprising nuclease activity or two engineered Cas9 systems comprising nickase activity, as well as the donor polynucleotide, such that the engineered Cas9 system or systems introduce a double-stranded break in the target site in the chromosomal sequence and repair of the double-stranded break by cellular DNA repair processes leads to insertion or exchange of sequence in the donor polynucleotide into the target site in the chromosomal sequence (i.e., gene correction or gene knock-in).

[0059] In embodiments, in which the engineered Cas9 protein comprises epigenetic modification activity or transcriptional regulation activity, the chromosomal sequence modification can comprise a conversion of at least one nucleotide in or near the target site, a modification of at least one nucleotide in or near the target site, a modification of at least one histone protein in or near the target site, and/or a change in transcription in or near the target site in the chromosomal sequence.

(a) Introduction into the Cell

[0060] As mentioned above, the method comprises introducing into the eukaryotic cell at least one engineered Cas9 system and/or nucleic acid encoding said system (and optional donor polynucleotide). The at least one system and/or nucleic acid/donor polynucleotide can be introduced into the cell of interest by a variety of means.

[0061] In some embodiments, the cell can be transfected with the appropriate molecules (*i.e.*, protein, DNA, and/or RNA). Suitable transfection methods include nucleofection (or electroporation), calcium phosphate-mediated transfection, cationic polymer transfection (*e.g.*, DEAE-dextran or polyethylenimine), viral transduction, virosome transfection, virion transfection, liposome transfection, cationic liposome transfection, immunoliposome transfection, nonliposomal lipid transfection, dendrimer transfection, heat shock transfection, magnetofection, lipofection, gene gun delivery, impalefection, sonoporation, optical transfection, and proprietary agent-enhanced uptake of nucleic acids. Transfection methods are well known in the art (*see*, *e.g.*, "Current Protocols in Molecular Biology" Ausubel *et al.*, John Wiley & Sons, New York, 2003 or "Molecular Cloning: A Laboratory Manual" Sambrook & Russell, Cold Spring Harbor Press, Cold Spring Harbor, NY, 3rd edition, 2001). In other embodiments, the molecules can be introduced into the cell by microinjection. For example, the molecules can be injected into the cytoplasm or nuclei of the cells of interest. The amount of each molecule introduced into the cell can vary, but those skilled in the art are familiar with means for determining the appropriate amount.

[0062] The various molecules can be introduced into the cell simultaneously or sequentially. For example, the engineered Cas9 system (or its encoding nucleic acid) and the donor polynucleotide can be introduced at the same time. Alternatively, one can be introduced first and then the other can be introduced later into the cell.

[0063] In general, the cell is maintained under conditions appropriate for cell growth and/or maintenance. Suitable cell culture conditions are well known in the art and are described, for example, in Santiago *et al.*, Proc. Natl. Acad. Sci. USA, 2008,

105:5809-5814; Moehle *et al*. Proc. Natl. Acad. Sci. USA, 2007, 104:3055-3060; Urnov *et al*., Nature, 2005, 435:646-651; and Lombardo *et al*., Nat. Biotechnol., 2007, 25:1298-1306. Those of skill in the art appreciate that methods for culturing cells are known in the art and can and will vary depending on the cell type. Routine optimization may be used, in all cases, to determine the best techniques for a particular cell type.

(b) Optional Donor Polynucleotide

[0064] In embodiments in which the engineered Cas9 protein comprises nuclease or nickase activity, the method can further comprise introducing at least one donor polynucleotide into the cell. The donor polynucleotide can be single-stranded or double-stranded, linear or circular, and/or RNA or DNA. In some embodiments, the donor polynucleotide can be a vector, *e.g.*, a plasmid vector.

[0065] The donor polynucleotide comprises at least one donor sequence. In some aspects, the donor sequence of the donor polynucleotide can be a modified version of an endogenous or native chromosomal sequence. For example, the donor sequence can be essentially identical to a portion of the chromosomal sequence at or near the sequence targeted by the engineered Cas9 system, but which comprises at least one nucleotide change. Thus, upon integration or exchange with the native sequence, the sequence at the targeted chromosomal location comprises at least one nucleotide change. For example, the change can be an insertion of one or more nucleotides, a deletion of one or more nucleotides, a substitution of one or more nucleotides, or combinations thereof. As a consequence of the "gene correction" integration of the modified sequence, the cell can produce a modified gene product from the targeted chromosomal sequence.

[0066] In other aspects, the donor sequence of the donor polynucleotide can be an exogenous sequence. As used herein, an "exogenous" sequence refers to a sequence that is not native to the cell, or a sequence whose native location is in a different location in the genome of the cell. For example, the exogenous sequence can comprise protein coding sequence, which can be operably linked to an exogenous promoter control sequence such that, upon integration into the genome, the cell is able

to express the protein coded by the integrated sequence. Alternatively, the exogenous sequence can be integrated into the chromosomal sequence such that its expression is regulated by an endogenous promoter control sequence. In other iterations, the exogenous sequence can be a transcriptional control sequence, another expression control sequence, an RNA coding sequence, and so forth. As noted above, integration of an exogenous sequence into a chromosomal sequence is termed a "knock in."

[0067] As can be appreciated by those skilled in the art, the length of the donor sequence can and will vary. For example, the donor sequence can vary in length from several nucleotides to hundreds of nucleotides to hundreds of thousands of nucleotides.

[0068] Typically, the donor sequence in the donor polynucleotide is flanked by an upstream sequence and a downstream sequence, which have substantial sequence identity to sequences located upstream and downstream, respectively, of the sequence targeted by the engineered Cas9 system. Because of these sequence similarities, the upstream and downstream sequences of the donor polynucleotide permit homologous recombination between the donor polynucleotide and the targeted chromosomal sequence such that the donor sequence can be integrated into (or exchanged with) the chromosomal sequence.

[0069] The upstream sequence, as used herein, refers to a nucleic acid sequence that shares substantial sequence identity with a chromosomal sequence upstream of the sequence targeted by the engineered Cas9 system. Similarly, the downstream sequence refers to a nucleic acid sequence that shares substantial sequence identity with a chromosomal sequence downstream of the sequence targeted by the engineered Cas9 system. As used herein, the phrase "substantial sequence identity" refers to sequences having at least about 75% sequence identity. Thus, the upstream and downstream sequences in the donor polynucleotide can have about 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with sequence upstream or downstream to the target sequence. In an exemplary embodiment, the upstream and downstream sequences in the donor polynucleotide can have about 95%

or 100% sequence identity with chromosomal sequences upstream or downstream to the sequence targeted by the engineered Cas9 system.

[0070] In some embodiments, the upstream sequence shares substantial sequence identity with a chromosomal sequence located immediately upstream of the sequence targeted by the engineered Cas9 system. In other embodiments, the upstream sequence shares substantial sequence identity with a chromosomal sequence that is located within about one hundred (100) nucleotides upstream from the target sequence. Thus, for example, the upstream sequence can share substantial sequence identity with a chromosomal sequence that is located about 1 to about 20, about 21 to about 40, about 41 to about 60, about 61 to about 80, or about 81 to about 100 nucleotides upstream from the target sequence. In some embodiments, the downstream sequence shares substantial sequence identity with a chromosomal sequence located immediately downstream of the sequence targeted by the engineered Cas9 system. In other embodiments, the downstream sequence shares substantial sequence identity with a chromosomal sequence that is located within about one hundred (100) nucleotides downstream from the target sequence. Thus, for example, the downstream sequence can share substantial sequence identity with a chromosomal sequence that is located about 1 to about 20, about 21 to about 40, about 41 to about 60, about 61 to about 80, or about 81 to about 100 nucleotides downstream from the target sequence.

[0071] Each upstream or downstream sequence can range in length from about 20 nucleotides to about 5000 nucleotides. In some embodiments, upstream and downstream sequences can comprise about 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2800, 3000, 3200, 3400, 3600, 3800, 4000, 4200, 4400, 4600, 4800, or 5000 nucleotides. In specific embodiments, upstream and downstream sequences can range in length from about 50 to about 1500 nucleotides.

(c) Cell Types

[0072] A variety of eukaryotic cells are suitable for use in the methods disclosed herein. For example, the cell can be a human cell, a non-human mammalian cell, a non-mammalian vertebrate cell, an invertebrate cell, an insect cell, a plant cell, a yeast cell, or a single cell eukaryotic organism. In some embodiments, the cell can be a one cell embryo. For example, a non-human mammalian embryo including rat, hamster, rodent, rabbit, feline, canine, ovine, porcine, bovine, equine, and primate embryos. In still other embodiments, the cell can be a stem cell such as embryonic stem cells, ES-like stem cells, fetal stem cells, adult stem cells, and the like. In one embodiment, the stem cell is not a human embryonic stem cell. Furthermore, the stem cells may include those made by the techniques disclosed in WO2003/046141, which is incorporated herein in its entirety, or Chung *et al.* (Cell Stem Cell, 2008, 2:113-117). The cell can be *in vitro* (*i.e.*, in culture), *ex vivo* (*i.e.*, within tissue isolated from an organism), or *in vivo* (*i.e.*, within an organism). In exemplary embodiments, the cell is a mammalian cell or mammalian cell line. In particular embodiments, the cell is a human cell or human cell line.

[0073] Non-limiting examples of suitable mammalian cells or cell lines include human embryonic kidney cells (HEK293, HEK293T); human cervical carcinoma cells (HELA); human lung cells (W138); human liver cells (Hep G2); human U2-OS osteosarcoma cells, human A549 cells, human A-431 cells, and human K562 cells; Chinese hamster ovary (CHO) cells, baby hamster kidney (BHK) cells; mouse myeloma NS0 cells, mouse embryonic fibroblast 3T3 cells (NIH3T3), mouse B lymphoma A20 cells; mouse melanoma B16 cells; mouse myoblast C2C12 cells; mouse myeloma SP2/0 cells; mouse embryonic mesenchymal C3H-10T1/2 cells; mouse carcinoma CT26 cells, mouse prostate DuCuP cells; mouse breast EMT6 cells; mouse hepatoma Hepa1c1c7 cells; mouse myeloma J5582 cells; mouse epithelial MTD-1A cells; mouse myocardial MyEnd cells; mouse renal RenCa cells; mouse pancreatic RIN-5F cells; mouse melanoma X64 cells; mouse lymphoma YAC-1 cells; rat glioblastoma 9L cells; rat B lymphoma RBL cells; rat neuroblastoma B35 cells; rat hepatoma cells (HTC); buffalo rat liver BRL 3A cells; canine kidney cells (MDCK); canine mammary (CMT)

cells; rat osteosarcoma D17 cells; rat monocyte/macrophage DH82 cells; monkey kidney SV-40 transformed fibroblast (COS7) cells; monkey kidney CVI-76 cells; African green monkey kidney (VERO-76) cells. An extensive list of mammalian cell lines may be found in the American Type Culture Collection catalog (ATCC, Manassas, VA).

(V) Applications

[0074] The compositions and methods disclosed herein can be used in a variety of therapeutic, diagnostic, industrial, and research applications. In some embodiments, the present disclosure can be used to modify any chromosomal sequence of interest in a cell, animal, or plant in order to model and/or study the function of genes, study genetic or epigenetic conditions of interest, or study biochemical pathways involved in various diseases or disorders. For example, transgenic organisms can be created that model diseases or disorders, wherein the expression of one or more nucleic acid sequences associated with a disease or disorder is altered. The disease model can be used to study the effects of mutations on the organism, study the development and/or progression of the disease, study the effect of a pharmaceutically active compound on the disease, and/or assess the efficacy of a potential gene therapy strategy.

[0075] In other embodiments, the compositions and methods can be used to perform efficient and cost effective functional genomic screens, which can be used to study the function of genes involved in a particular biological process and how any alteration in gene expression can affect the biological process, or to perform saturating or deep scanning mutagenesis of genomic loci in conjunction with a cellular phenotype. Saturating or deep scanning mutagenesis can be used to determine critical minimal features and discrete vulnerabilities of functional elements required for gene expression, drug resistance, and reversal of disease, for example.

[0076] In further embodiments, the compositions and methods disclosed herein can be used for diagnostic tests to establish the presence of a disease or disorder and/or for use in determining treatment options. Examples of suitable diagnostic tests include detection of specific mutations in cancer cells (*e.g.*, specific

mutation in EGFR, HER2, and the like), detection of specific mutations associated with particular diseases (*e.g.*, trinucleotide repeats, mutations in β-globin associated with sickle cell disease, specific SNPs, *etc.*), detection of hepatitis, detection of viruses (*e.g.*, Zika), and so forth.

[0077] In additional embodiments, the compositions and methods disclosed herein can be used to correct genetic mutations associated with a particular disease or disorder such as, *e.g.*, correct globin gene mutations associated with sickle cell disease or thalassemia, correct mutations in the adenosine deaminase gene associated with severe combined immune deficiency (SCID), reduce the expression of HTT, the disease-causing gene of Huntington's disease, or correct mutations in the rhodopsin gene for the treatment of retinitis pigmentosa. Such modifications may be made in cells *ex vivo*.

[0078] In still other embodiments, the compositions and methods disclosed herein can be used to generate crop plants with improved traits or increased resistance to environmental stresses. The present disclosure can also be used to generate farm animal with improved traits or production animals. For example, pigs have many features that make them attractive as biomedical models, especially in regenerative medicine or xenotransplantation.

DEFINITIONS

[0079] Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by a person skilled in the art to which this invention belongs. The following references provide one of skill with a general definition of many of the terms used in this invention: Singleton *et al.*, Dictionary of Microbiology and Molecular Biology (2nd Ed. 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988); The Glossary of Genetics, 5th Ed., R. Rieger *et al.* (eds.), Springer Verlag (1991); and Hale & Marham, The Harper Collins Dictionary of Biology (1991). As used herein, the following terms have the meanings ascribed to them unless specified otherwise.

[0080] When introducing elements of the present disclosure or the preferred embodiments(s) thereof, the articles “a”, “an”, “the” and “said" are intended to mean that there are one or more of the elements. The terms “comprising”, “including” and “having” are intended to be inclusive and mean that there may be additional elements other than the listed elements.

[0081] The term “about” when used in relation to a numerical value, x, for example means x ± 5%.

[0082] As used herein, the terms “complementary” or “complementarity” refer to the association of double-stranded nucleic acids by base pairing through specific hydrogen bonds. The base paring may be standard Watson-Crick base pairing (*e.g.*, 5’-A G T C-3’ pairs with the complementary sequence 3’-T C A G-5’). The base pairing also may be Hoogsteen or reversed Hoogsteen hydrogen bonding. Complementarity is typically measured with respect to a duplex region and thus, excludes overhangs, for example. Complementarity between two strands of the duplex region may be partial and expressed as a percentage (*e.g.*, 70%), if only some (*e.g.*, 70%) of the bases are complementary. The bases that are not complementary are “mismatched.” Complementarity may also be complete (*i.e.*, 100%), if all the bases in the duplex region are complementary.

[0083] As used herein, the term “CRISPR/Cas system” or “Cas9 system” refers to a complex comprising a Cas9 protein (*i.e.*, nuclease, nickase, or catalytically dead protein) and a guide RNA.

[0084] The term “endogenous sequence,” as used herein, refers to a chromosomal sequence that is native to the cell.

[0085] As used herein, the term “exogenous” refers to a sequence that is not native to the cell, or a chromosomal sequence whose native location in the genome of the cell is in a different chromosomal location.

[0086] A “gene,” as used herein, refers to a DNA region (including exons and introns) encoding a gene product, as well as all DNA regions which regulate the production of the gene product, whether or not such regulatory sequences are adjacent to coding and/or transcribed sequences. Accordingly, a gene includes, but is not

necessarily limited to, promoter sequences, terminators, translational regulatory sequences such as ribosome binding sites and internal ribosome entry sites, enhancers, silencers, insulators, boundary elements, replication origins, matrix attachment sites, and locus control regions.

[0087] The term "heterologous" refers to an entity that is not endogenous or native to the cell of interest. For example, a heterologous protein refers to a protein that is derived from or was originally derived from an exogenous source, such as an exogenously introduced nucleic acid sequence. In some instances, the heterologous protein is not normally produced by the cell of interest.

[0088] The term "nickase" refers to an enzyme that cleaves one strand of a double-stranded nucleic acid sequence (*i.e.*, nicks a double-stranded sequence). For example, a nuclease with double strand cleavage activity can be modified by mutation and/or deletion to function as a nickase and cleave only one strand of a double-stranded sequence.

[0089] The term "nuclease," as used herein, refers to an enzyme that cleaves both strands of a double-stranded nucleic acid sequence.

[0090] The terms "nucleic acid" and "polynucleotide" refer to a deoxyribonucleotide or ribonucleotide polymer, in linear or circular conformation, and in either single- or double-stranded form. For the purposes of the present disclosure, these terms are not to be construed as limiting with respect to the length of a polymer. The terms can encompass known analogs of natural nucleotides, as well as nucleotides that are modified in the base, sugar and/or phosphate moieties (*e.g.*, phosphorothioate backbones). In general, an analog of a particular nucleotide has the same base-pairing specificity; *i.e.*, an analog of A will base-pair with T.

[0091] The term "nucleotide" refers to deoxyribonucleotides or ribonucleotides. The nucleotides may be standard nucleotides (*i.e.*, adenosine, guanosine, cytidine, thymidine, and uridine), nucleotide isomers, or nucleotide analogs. A nucleotide analog refers to a nucleotide having a modified purine or pyrimidine base or a modified ribose moiety. A nucleotide analog may be a naturally occurring nucleotide (*e.g.*, inosine, pseudouridine, etc.) or a non-naturally occurring nucleotide.

Non-limiting examples of modifications on the sugar or base moieties of a nucleotide include the addition (or removal) of acetyl groups, amino groups, carboxyl groups, carboxymethyl groups, hydroxyl groups, methyl groups, phosphoryl groups, and thiol groups, as well as the substitution of the carbon and nitrogen atoms of the bases with other atoms (*e.g.*, 7-deaza purines). Nucleotide analogs also include dideoxy nucleotides, 2'-O-methyl nucleotides, locked nucleic acids (LNA), peptide nucleic acids (PNA), and morpholinos.

[0092] The terms "polypeptide" and "protein" are used interchangeably to refer to a polymer of amino acid residues.

[0093] The terms "target sequence," "target chromosomal sequence," and "target site" are used interchangeably to refer to the specific sequence in chromosomal DNA to which the engineered Cas9 system is targeted, and the site at which the engineered Cas9 system modifies the DNA or protein(s) associated with the DNA.

[0094] Techniques for determining nucleic acid and amino acid sequence identity are known in the art. Typically, such techniques include determining the nucleotide sequence of the mRNA for a gene and/or determining the amino acid sequence encoded thereby, and comparing these sequences to a second nucleotide or amino acid sequence. Genomic sequences can also be determined and compared in this fashion. In general, identity refers to an exact nucleotide-to-nucleotide or amino acid-to-amino acid correspondence of two polynucleotides or polypeptide sequences, respectively. Two or more sequences (polynucleotide or amino acid) can be compared by determining their percent identity. The percent identity of two sequences, whether nucleic acid or amino acid sequences, is the number of exact matches between two aligned sequences divided by the length of the shorter sequences and multiplied by 100. An approximate alignment for nucleic acid sequences is provided by the local homology algorithm of Smith and Waterman, Advances in Applied Mathematics 2:482-489 (1981). This algorithm can be applied to amino acid sequences by using the scoring matrix developed by Dayhoff, Atlas of Protein Sequences and Structure, M. O. Dayhoff ed., 5 suppl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., USA, and normalized by Gribskov, Nucl. Acids Res. 14(6):6745-6763

(1986). An exemplary implementation of this algorithm to determine percent identity of a sequence is provided by the Genetics Computer Group (Madison, Wis.) in the "BestFit" utility application. Other suitable programs for calculating the percent identity or similarity between sequences are generally known in the art, for example, another alignment program is BLAST, used with default parameters. For example, BLASTN and BLASTP can be used using the following default parameters: genetic code=standard; filter=none; strand=both; cutoff=60; expect=10; Matrix=BLOSUM62; Descriptions=50 sequences; sort by=HIGH SCORE; Databases=non-redundant, GenBank+EMBL+DDBJ+PDB+GenBank CDS translations+Swiss protein+Spupdate+PIR. Details of these programs can be found on the GenBank website.

[0095] As various changes could be made in the above-described cells and methods without departing from the scope of the invention, it is intended that all matter contained in the above description and in the examples given below, shall be interpreted as illustrative and not in a limiting sense.

EXAMPLES

[0096] The following examples illustrate certain aspects of the disclosure.

Example 1: Determination of PAM requirements for target DNA cleavage by Cas9 orthologs

[0097] Cas9 orthologs from *Bacillus smithii*, *Lactobacillus rhamnosus*, *Parasutterella excrementihominis*, *Mycoplasma canis*, *Mycoplasma gallisepticum*, *Akkermansia glycaniphila*, *Akkermansia muciniphila*, *Oenococcus kitaharae*, *Bifidobacterium bombi*, *Acidothermus cellulolyticus*, *Alicyclobacillus hesperidum*, *Wolinella succinogenes*, *Nitratifractor salsuginis*, *Ralstonia syzygii*, and *Corynebacterium diphtheria* were codon optimized for expression in human cells and tagged with a SV40 large T antigen nuclear localization (NLS) on the C terminus (SEQ ID NOs:1-30; see Table 6 below). The expression of each ortholog was driven by a human cytomegalovirus (CMV) immediate early enhancer and promoter. CRISPR RNA

(crRNA) and putative trans-activating crRNA (tracrRNA) for each ortholog were joined together to form a single guide RNA (sgRNA) (SEQ ID NOs:31-45; see Table 6 below). The expression of each sgRNA was driven by a human U6 promoter. *In vitro* transcribed sgRNA was prepared from a T7 promoter tagged PCR template as a supplement for *in vitro* digestion.

[0098] Human K562 cells were transfected with Cas9 encoding plasmid and sgRNA expression plasmid by nucleofection. Each transfection consisted of 2 million cells, 5 µg of Cas9 encoding plasmid DNA, and 3 µg of sgRNA expression plasmid DNA. Cells were harvested approximately 24 hr post transfection, washed with ice cold PBS buffer, and lysed with 150 µL of lysis solution (20 mM HEPES, pH 7.5; 100 mM KCl; 5 mM MgCl2, 1 mM DTT, 5% glycerol, 0.1% Triton X-100, 1x Protease inhibitor) with constant agitation for 30 minutes in a 4 °C cold room. Supernatant was prepared by removing residual cellular debris with centrifugation at 16,000 x g for 2 minutes at 4 °C and used as a source of Cas9 RNP for *in vitro* digestion of a plasmid DNA PAM library. The library contained 4^8 degenerate PAMs, each immediately preceded by a protospacer with the following configuration: 5'-GTACAAACGGCAGAAGCTGGNNNNNNNN-3' (SEQ ID NO:46). Each *in vitro* digestion consisted of 10 µL of cell lysate supernatant, 2 µL of 5x digestion buffer (100 mM HEPES, pH 7.5; 500 mM KCl; 25 mM $MgCl_2$; 5 mM DTT; 25% glycerol), 800 ng of PAM library DNA, and 20 pmol of *in vitro* transcribed sgRNA supplement in a 20 µL reaction volume. Reaction was maintained at 37 °C for 30 minutes and then purified with PCR purification kit. Illumina NextSeq sequencing libraries were prepared from digested products and subjected to deep sequencing. Deep sequencing data were analyzed using a Weblogo program to deduce the PAM requirement for each Cas9 ortholog.

[0099] Results are summarized in FIG. **1**. The results revealed several Cas9 orthologs that use A and/or T containing PAMs for *in vitro* target DNA cleavage. These Cas9 orthologs could provide a means to target AT rich genomic sites. The results also revealed several Cas9 orthologs that use a PAM suitable for targeting GC

rich genomic sites. These Cas9 orthologs could provide alternative targeting schemes to SpyCas9 in GC rich genomic sites to increase targeting resolution and specificity.

Example 2: Genome modification using *Bacillus smithii* Cas9 (BsmCas9) and *Lactobacillus rhamnosus* Cas9 (LrhCas9)

[0100] As shown in FIG. **1** and Table A (above), the small BsmCas9 (1095 aa) (SEQ ID NO: 2) and the LrhCas9 (SEQ ID NO: 4) use a 5'-NNNNCAAA-3' PAM and a 5'-NGAAA-3' PAM for target DNA binding, respectively. These novel PAM usages provide a means to target AT rich genomic sites. To demonstrate gene editing, human K562 cells ($1x10^6$) were nucleofected with 5 µg of Cas9 encoding plasmid DNA and 3 µg of sgRNA expression plasmid DNA. Targeted genomic sites include the human tyrosine-protein phosphatase non-receptor type 2 (*PTN2*) locus, the human empty spiracles homeobox 1 (*EMX1*) locus, the human programmed cell death 1 ligand 1 (*PD1L1*) locus, the human AAVS1 safe harbor locus, the human cytochrome p450 oxidoreductase (*POR*) locus, and the human nuclear receptor subfamily 1 group I member 3(*CAR*) locus. Genomic DNA was prepared using a DNA extraction solution (QuickExtract™) three days post transfection and targeted genomic regions were each PCR amplified (JumpStart Taq™ ReadyMix™). The PCR primers are listed in Table 1.

Table 1. PCR Primers.

Locus	Forward primer (5'-3')	Reverse primer (5'-3')	Size (bp)
PTN2	CTGTTTCCTGGGTTCCAATAA CAAGAC (SEQ ID NO:47)	ACAAGGGCTCAAGTGGAGTG (SEQ ID NO:48)	290
EMX1	ATGGGAGCAGCTGGTCAGAG (SEQ ID NO:49)	CAGCCCATTGCTTGTCCCT (SEQ ID NO:50)	507
PD1L1	CTCGCCATTCCAGCCACTCA AAC (SEQ ID NO:51)	GGTTAAGTCGGGTTTCCTTG CAG (SEQ ID NO:52)	341
AAVS1	TTCGGGTCACCTCTCACTCC (SEQ ID NO:53)	GGCTCCATCGTAAGCAAACC (SEQ ID NO:54)	469

POR	CTCCCCTGCTTCTTGTCGTAT (SEQ ID NO:55)	ACAGGTCGTGGACACTCACA (SEQ ID NO:56)	380
CAR	GGATCAAGTCAAGGGCATGT (SEQ ID NO:57)	ATGTAGCTGGACAGGCTTGG (SEQ ID NO:58)	347

[0101] Amplification was carried out using the following condition: 1 cycle of 98 °C for 2 minutes for initial denaturation; 34 cycles of 98 °C for 15 seconds, 62 °C for 30 seconds, and 72 °C for 45 seconds; 1 cycle of 72 °C for 5 minutes; and hold at 4 °C. PCR products were digested with Cel-1 nuclease and resolved on a 10% acrylamide gel. Targeted mutation rates were measured using ImageJ and expressed as percent insertions and/or deletions (% Indel). Results are summarized in Table 2. These results demonstrate that both Cas9 orthologs were able to edit endogenous genomic sites in human cells using a 5'-NNNNCAAA-3' PAM (BsmCas9) or a 5'-NGAAA-3' PAM (LrhCas9).

Table 2. Gene Editing with BsmCas9 and LrhCas9 in Human K562 Cells.

Cas9	Locus/Target #	Protospacer sequence (5'-3')	PAM (5'-3')*	Indel (%)
BsmCas9	*PTN2*	CTCATACATGGCTATAATA GAA (SEQ ID NO:59)	GGAGCAAA	11.9
	EMX1/Target 1	GAAGGTGTGGTTCCAGAA CCGG (SEQ ID NO:60)	AGGACAAA	6.8
	EMX1/Target 2	TGGTTCCAGAACCGGAGG ACAA (SEQ ID NO:61)	AGTACAAA	10.9
	EMX1/Target 3	CCCAGGTGAAGGTGTGGT TCCA (SEQ ID NO:62)	GAACCGGA	0
	EMX1/Target 4	AGAACCGGAGGACAAAGT ACAA (SEQ ID NO:63)	ACGGCAGA	0

LrhCas9	*PD1L1*	CCTCTGGCACATCCTCCAA A (SEQ ID NO:64)	TGAAA	38.9
	AAVS1	CTAGGGACAGGATTGGTG AC (SEQ ID NO:65)	AGAAA	32.7
	POR/Target 1	GCTCGTACTGGCGAATGCT (SEQ ID NO:66)	GGAAA	26.7
	POR/Target 2	GCTGAAGAGCTACGAGAA CC (SEQ ID NO:67)	AGAAG	0
	POR/Target 3	CATGGGGGAGATGGGCCG GC (SEQ ID NO:68)	TGAAG	0
	CAR/Target 1	AGAGACTCTCTAGAAGGGA C (SEQ ID NO:69)	AGAAA	31.7
	CAR/Target 2	GTGAGAGTCTCCTCCCCAA TG (SEQ ID NO:70)	GGAAA	27.0
	CAR/Target 3	GGGAGGAGACTCTCACCT GA (SEQ ID NO:71)	AGAAA	0
* The determinant nucleotides of the PAM are underlined.				

Example 3: Improvement of *Parasutterella excrementihominis* Cas9 (PexCas9) by fusion with chromatin modulating motifs

[0102] *Parasutterella excrementihominis* Cas9 (PexCas9-NLS) (SEQ ID NO:6) was modified by fusion with a human HMGN1 peptide (SEQ ID NO:72) on the N terminus using a TGSG linker (SEQ ID NO:109) and with either a human HMGB1 box A peptide (PexCas9-HN1HB1 fusion; SEQ ID NO:117) or a human histone H1 central globular domain peptide (PexCas9-HN1H1G; SEQ ID NO:118) on the C terminus using a LEGGGS linker (SEQ ID NO:108).

Table 3. Chromatin Modulating Motifs

Name	Peptide Sequence	SEQ ID NO:
HMGN1 (HN1)	MPKRKVSSAEGAAKEEPKRRSARLSAKPPAKV EAKPKKAAAKDKSSDKKVQTKGKRGAKGKQAE VANQETKEDLPAENGETKTEESPASDEAGEKEA KSD	72
Human HMGB1 box A (HB1)	MGKGDPKKPRGKMSSYAFFVQTCREEHKKKHP DASVNFSEFSKKCSERWKTMSAKEKGKFEDMA KADKARYEREMKTYIPPKGE	73
Human histone H1 central globular domain (H1G)	STDHPKYSDMIVAAIQAEKNRAGSSRQSIQKYIK SHYKVGENADSQIKLSIKRLVTTGVLKQTKGVGA SGSFRLAKSDEP	74

[0103] Human K562 cells (1x 10^6) were transfected with plasmid DNA encoding PexCas9-NLS, PexCas9-HN1HB1 fusion, or PexCas9-HN1H1G fusion in molar equivalent amounts (5 and 5.4 µg, respectively) and 3 µg of sgRNA plasmid for targeting a genomic site in the human cytochrome p450 oxidoreductase (*POR*) locus. Genomic DNA was prepared using DNA extraction solution (QuickExtract™) three days post transfection and the targeted genomic region was PCR amplified using the forward primer 5'- CTCCCCTGCTTCTTGTCGTAT-3' (SEQ ID NO:55) and the reverse primer 5'- ACAGGTCGTGGACACTCACA -3' (SEQ ID NO:56). Amplification was carried out with the following condition: 1 cycle of 98 °C for 2 minutes for initial denaturation; 34 cycles of 98 °C for 15 seconds, 62 °C for 30 seconds, and 72 °C for 45 seconds; 1 cycle of 72 °C for 5 minutes; and hold at 4 °C. PCR products were digested with Cel-1 nuclease and resolved on a 10% acrylamide gel. Targeted mutation rates were measured using ImageJ and expressed as percent insertions and/or deletions (% Indel). Results are summarized in Table 4. The results demonstrate that Cas9 fusion with at least one chromatin modulating motif enhances its gene editing efficiency on endogenous targets in human cells.

Table 4. Gene editing using PexCas9 and PexCas9 fusion proteins in human K562 cells.

Cas9 nuclease	Locus	Protospacer (5'-3')	PAM (5'-3')*	Indel (%)
PexCas9-NLS	*POR*	TGTACATGGGGGAGATGGGC (SEQ ID NO:75)	CGG	22.9
PexCas9-HN1HB1 fusion				36.8
PexCas9HN1H1G fusion				43.7

* The determinant nucleotides of the PAM are underlined.

Example 4. Improvement of *Mycoplasma canis* Cas9 (McaCas9) system by sgRNA modification

[0104] The wild type crRNA coding sequence of McaCas9 contains four consecutive thymidine residues in the repeat region, and three of the four thymidine residues are predicted to pair with three adenosine residues in the putative tracrRNA sequence when the crRNA and tracrRNA are joined together to form a sgRNA. Human RNA polymerase (Pol) III is known to use four or more consecutive thymidine residues on the coding RNA strand as a transcription termination signal. To prevent an early transcriptional termination of McaCas9 sgRNA in human cells, a T to C mutation and a corresponding A to G mutation were introduced into the sgRNA scaffold to form a modified sgRNA scaffold with the following sequence: 5'-GUUCUAGUGUUGUACAAUAUUUGGGUGAAAACCCAAAUAUUGUACAUCCUAGAUCAAGGCGCUUAAUUGCUGCCGUAAUUGCUGAAAGCGUAGCUUUCAGUUUUUUU-3' (SEQ ID NO:76), where the mutated nucleotides are underlined. This modification is also predicted to increase the sgRNA scaffold thermodynamic stability.

[0105] Human K562 cells (1x 10^6) were transfected with 5.5 µg of plasmid DNA encoding a McaCas9 fusion protein, which contains a HMGN1 peptide on the N terminus and a histone H1 globular domain peptide on the C terminus, and 3 µg of sgRNA plasmid DNA encoding the control sgRNA scaffold or the modified sgRNA

scaffold. Genomic DNA was prepared using a DNA extraction solution (QuickExtract™) three days post transfection and the targeted genomic region was PCR amplified using the forward primer 5'- CTCCCCTGCTTCTTGTCGTAT-3' (SEQ ID NO:55) and the reverse primer 5'- ACAGGTCGTGGACACTCACA -3' (SEQ ID NO:56). Amplification was carried out with the following condition: 1 cycle of 98 °C for 2 minutes for initial denaturation; 34 cycles of 98 °C for 15 seconds, 62 °C for 30 seconds, and 72 °C for 45 seconds; 1 cycle of 72 °C for 5 minutes; and hold at 4 °C. PCR products were digested with Cel-1 nuclease and resolved on a 10% acrylamide gel. Targeted mutation rates were measured using ImageJ and expressed as percent insertions and/or deletions (% Indel). Results are summarized in Table 5. The results demonstrate that the activity of a Cas9 ortholog in mammalian cells can be enhanced by modifying its sgRNA scaffold.

Table 5. Gene editing using McaCas9 in combination with control sgRNA scaffold or modified sgRNA scaffold.

sgRNA scaffold	Locus	Protospacer (5'-3')	PAM (5'-3')*	Indel (%)
Control sgRNA scaffold	*POR*	ATAGATGCGGCCAAGGTGTACA (SEQ ID NO:77)	TGGG	25.5
Modified sgRNA scaffold				36.2

* The determinant nucleotides of the PAM are underlined.

Example 5. Improvement of McaCas9, BsmCas9, PexCas9, and LrhCas9 activity by fusion with chromatin modulating motifs

[0106] Additional Cas9-CMM fusion proteins were prepared by linking McaCas9-NLS, BsmCas9-NLS, and LrhCas9-NLS proteins with HMGN1 (HN1) at the amino terminus and either HMGB1 box A (HB1) or histone H1 central globular motif (H1G) at the carboxyl terminus to yield McaCas9-HN1HB1 (SEQ ID NO:123), McaCas9-HN1H1G (SEQ ID NO:124), BsmCas9-HN1HB1 (SEQ ID NO:119), Bsm-HN1H1G (SEQ ID NO:120), Lrh-HN1HB1 (SEQ ID NO:121), LrhCas9-HN1H1G (SEQ

ID NO:122). The nuclease activity of these fusions and the PexCas9-CMM fusions described above in Example 3 were compared to the activity of the corresponding engineered Cas9 protein essentially as described above in Examples 2 and 3. Table 6 presents the target site (*i.e.*, protospacer + PAM, which is shown in bold with the determinate nucleotides underlined) in specific loci for each Cas9 nuclease.

Table 6. Gene Editing Target Sites

Cas9	Locus	Target site (5'-3')	SEQ ID NO
Mca	*POR1*	ATAGATGCGGCCAAGGTGTACA**TGGG**	125
	POR2	CTACGAGAACCAGAAGCCGTGA**GTGG**	126
Bsm	*PTN2*	CTCATACATGGCTATAATAGAA**GGAGCAAA**	127
	EMX1	GAAGGTGTGGTTCCAGAACCGG**AGGACAAA**	128
	EMX2	TGGTTCCAGAACCGGAGGACAA**AGTACAAA**	129
Pex	*POR1*	TGTACATGGGGGAGATGGGC**CGG**	130
	AAVS1	GGGGCCACTAGGGACAGGAT**TGG**	131
Lrh	*POR1*	AGCTCGTACTGGCGAATGCT**GGAAA**	132
	PD1L1	CCTCTGGCACATCCTCCAAA**TGAAA**	133

[0107] The percent of indels under each condition are plotted in FIGS. **2A-D**. Both the HN1HB1 and HN1H1G combinations significantly enhanced the four Cas9 orthologs on at least one site. Based on the fold change magnitude, the CMM fusion modification provided the largest enhancement on McaCas9, increasing its activity by at least five-fold on the two sites tested (FIG. **2A**). CMM fusion provided more than two-fold enhancement on PexCas9 (FIG. 2B). BsmCas9 activity was enhanced by more than three-fold on one site, but there was only 20% increase on the second site and no effect on the third site (FIG. **2C**). It should be noted, however, that all three BsmCas9 nucleases were highly efficient (> 35% indels). LrhCas9 was highly efficient on the two sites tested (22% and 33% indels) even without the fusion modification (FIG. **2D**). However, the HN1H1G combination still provided a significant enhancement on both

sites, with 70% and 28% increase in activity. These results demonstrate that the CMM fusion strategy enhances gene editing efficiency.

Example 6. Off-target effects of Cas9-CMM fusions

[0108] To assess the off-target activity of the Cas9-CMM fusions, 1 to 5 top ranked potential off-target sites for each target site were analyzed using the Surveyor Nuclease assay. In addition to the Cas9 and Cas9-CMM fusion data described above in Example 5, data from *Streptococcus pyogenes* Cas9 (SpyCas9), SpyCas9-CMM fusions, *Streptococcus pasteurianus* Cas9 (SpaCas9), Spa-CMM fusions, *Campylobacter jejuni* Cas9 (CjeCas9), and CjeCas9-SMM fusions were also analyzed. From a total of 64 potential off-target sites assayed, off-target cleavage was detected on 11 sites, contributed by 9 guide sequences out of total 21 guide sequences tested. On the 11 off-target sites, the control Cas9 and the fusion nucleases were concurrent, with the exception of the POR Spy 1-OT1 site, where no off-target cleavage was detected on the control SpyCas9. Overall, there was no significant difference between the fusion nucleases and the control Cas9 (FIG. **3**). For example, across all the 11 off-target sites, the HN1H1G fusion combination averaged 8.0 ± 6.0% indels and the control Cas9 averaged 7.5 ± 5.1% indels. Likewise, across the 10 off-target sites relevant to the HN1HB1 fusion combination, there was no significant difference between the fusion combination and the control Cas9 (6.9 ± 5.7% vs. 6.5 ± 5.4% indels). Taken together, these results show that the on-target activity enhancement by the HN1H1B and HN1H1G fusion combinations generally does not result in an increase in off-target activity.

Engineered Cas9 Systems

[0109] Table 7 presents the human codon optimized DNA and protein sequences of engineered Cas9/NLS proteins (SEQ ID NOS:1-30, wherein the NLS sequence is underlined) and the DNA sequences of engineered sgRNAs (SEQ ID NOS:31-45; the N residues at the 5' end indicate the programmable target sequence). Also presented at the Cas9-CMM fusions (SEQ ID NOS:117-124).

Table 7. Engineered Cas9 Systems

BsmCas9/NLS DNA sequence (SEQ ID NO:1)

ATGAACTACAAGATGGGCCTCGACATCGGAATCGCCTCTGTTGGATGGGCCGTGATCAACCTG
GACCTGAAGAGAATCGAGGACCTCGGCGTGCGGATCTTCGACAAGGCTGAGCATCCTCAGAAC
GGCGAGTCTCTGGCCCTGCCTAGAAGAATTGCCAGAAGCGCCAGACGGCGGCTGCGGAGAAGA
AAGCACAGACTGGAACGGATCAGACGGCTGCTGGTGTCCGAGAACGTGCTGACCAAAGAAGAG
ATGAACCTGCTGTTCAAGCAGAAAAAGCAGATCGACGTGTGGCAGCTGAGAGTGGACGCCCTG
GAAAGAAAGCTGAACAACGACGAGCTGGCCAGAGTGCTGCTGCACCTGGCCAAGAGAAGAGGC
TTCAAGAGCAACAGAAAGAGCGAGCGGAACAGCAAAGAGAGCAGCGAGTTCCTGAAGAACATC
GAAGAGAACCAGAGCATTCTGGCCCAGTACAGATCCGTGGGCGAGATGATCGTGAAGGACAGC
AAGTTCGCCTACCACAAGCGGAACAAGCTGGACAGCTACAGCAACATGATCGCCAGGGACGAT
CTGGAAAGAGAGATCAAGCTGATCTTCGAGAAGCAGCGCGAGTTCAACAACCCCGTGTGCACC
GAGAGACTGGAAGAGAAGTACCTGAACATCTGGTCCAGCCAGCGGCCTTTCGCCTCCAAAGAG
GACATCGAGAAAAAAGTGGGCTTCTGCACCTTCGAGCCCAAAGAGAAAAGAGCCCCTAAGGCC
ACCTACACCTTCCAGAGCTTCATCGTGTGGGAGCACATCAACAAGCTGCGGCTGGTGTCTCCC
GACGAGACAAGAGCCCTGACCGAGATCGAGCGGAATCTGCTGTATAAGCAGGCCTTCAGCAAG
AACAAGATGACCTACTACGACATCCGGAAGCTGCTGAACCTGAGCGACGACATCCACTTCAAG
GGCCTGCTGTACGACCCCAAGAGCAGCCTGAAGCAGATTGAGAACATCCGGTTTCTGGAACTG
GACTCTTACCACAAGATCCGGAAGTGCATCGAGAATGTGTACGGCAAGGACGGCATCCGCATG
TTCAACGAGACAGACATCGACACCTTCGGCTACGCCCTGACCATCTTCAAGGACGACGAGGAT
ATCGTGGCCTACCTGCAGAACGAGTACATCACCAAGAACGGCAAGCGGGTGTCCAATCTGGCC
AACAAGGTGTACGACAAGTCCCTGATCGACGAACTGCTGAATCTGTCCTTCTCCAAATTCGCC
CACCTGAGCATGAAGGCCATCCGGAACATCCTGCCTTACATGGAACAGGGCGAAATCTACAGC
AAGGCCTGCGAACTGGCCGGCTACAACTTCACAGGCCCCAAGAAGAAAGAGAAGGCCCTGCTG
CTGCCTGTGATCCCCAATATCGCCAATCCTGTGGTCATGCGGGCCCTGACACAGAGCAGAAAG
GTGGTCAACGCCATCATCAAGAAATACGGATCCCCCGTGTCCATCCACATCGAGCTGGCTAGG
GATCTGAGCCACAGCTTCGACGAGCGGAAGAAGATCCAGAAGGACCAGACCGAGAACCGCAAG
AAGAACGAAACCGCCATCAAGCAGCTGATCGAGTACGAGCTGACTAAGAACCCCACCGGCCTG
GACATCGTGAAGTTCAAACTTTGGAGCGAGCAGCAAGGCAGATGCATGTACTCCCTGAAGCCT
ATTGAGCTGGAAAGACTGCTGGAACCCGGCTACGTGGAAGTGGACCACATTCTGCCCTACAGC
AGAAGCCTGGACGACAGCTACGCCAACAAAGTGCTGGTCCTGACAAAAGAGAACCGCGAAAAG
GGCAATCACACCCCTGTGGAATATCTCGGCCTGGGCTCTGAGCGGTGGAAGAAATTCGAGAAG
TTCGTGCTGGCTAACAAGCAGTTCTCTAAGAAGAAGAAGCAGAACCTGCTCCGGCTGAGATAC
GAGGAAACCGAGGAAAAAGAGTTCAAAGAGCGGAACCTGAACGACACCCGGTACATCTCCAAG
TTCTTCGCCAACTTCATCAAAGAGCATCTGAAGTTCGCCGACGGCGACGGCGGCCAGAAAGTG
TACACAATCAACGGCAAGATCACCGCTCACCTGAGAAGCAGATGGGACTTCAACAAGAACCGG
GAAGAGAGCGACCTGCACCACGCTGTGGATGCTGTGATTGTGGCCTGTGCCACACAGGGCATG
ATCAAGAAGATTACCGAGTTCTACAAGGCCCGCGAGCAGAACAAAGAGTCCGCCAAGAAAAAA
GAACCCATCTTTCCCCAGCCTTGGCCTCACTTCGCCGATGAGCTGAAGGCTCGGCTGAGCAAG
TTCCCTCAAGAGTCCATCGAGGCCTTCGCTCTGGGCAACTACGACAGAAAGAAGCTGGAATCC
CTGCGGCCTGTGTTCGTGTCCAGAATGCCCAAGAGATCCGTGACAGGCGCTGCCCACCAAGAG
ACACTGAGAAGATGCGTGGGCATCGACGAGCAGTCTGGCAAGATTCAGACCGCCGTGAAAACA
AAGCTGAGCGACATCAAGCTGGATAAGGACGGACACTTCCCCATGTACCAGAAAGAGTCTGAC
CCCAGAACCTACGAGGCCATCAGACAGAGGCTGCTCGAACACAACAACGACCCTAAGAAGGCC

TTTCAAGAGCCACTGTACAAGCCCAAAAAGAATGGCGAGCCCGGACCAGTGATCCGGACCGTG
AAGATCATCGACACAAAGAACAAGGTGGTGCACCTGGACGGCAGCAAGACAGTGGCCTACAAC
TCCAACATCGTGCGGACCGACGTGTTCGAGAAGGATGGCAAGTACTACTGCGTGCCCGTGTAC
ACTATGGATATCATGAAGGGCACCCTGCCTAACAAGGCCATCGAAGCCAACAAGCCCTACTCC
GAGTGGAAAGAGATGACCGAAGAGTACACGTTCCAGTTCAGTCTGTTCCCCAACGACCTCGTG
CGCATCGTGCTGCCAAGAGAGAAAACCATCAAGACCAGCACCAACGAGGAAATCATCATTAAG
GACATCTTTGCCTACTACAAGACCATCGACAGCGCCACAGGCGGCCTGGAACTGATCTCCCAC
GATCGGAACTTCAGCCTGAGAGGCGTGGGCTCTAAGACACTGAAGCGCTTTGAGAAGTATCAG
GTGGACGTGCTGGGCAACATCCACAAAGTGAAGGGCGAGAAGAGAGTCGGCCTGGCCGCTCCT
ACCAACCAGAAAAAGGGAAAGACCGTGGACAGCCTGCAGAGCGTGTCCGATCCCAAGAAGAAG
AGGAAGGTG

BsmCas9/NLS protein sequence (SEQ ID NO:2)

MNYKMGLDIGIASVGWAVINLDLKRIEDLGVRIFDKAEHPQNGESLALPRRIARSARRRLRRR
KHRLERIRRLLVSENVLTKEEMNLLFKQKKQIDVWQLRVDALERKLNNDELARVLLHLAKRRG
FKSNRKSERNSKESSEFLKNIEENQSILAQYRSVGEMIVKDSKFAYHKRNKLDSYSNMIARDD
LEREIKLIFEKQREFNNPVCTERLEEKYLNIWSSQRPFASKEDIEKKVGFCTFEPKEKRAPKA
TYTFQSFIVWEHINKLRLVSPDETRALTEIERNLLYKQAFSKNKMTYYDIRKLLNLSDDIHFK
GLLYDPKSSLKQIENIRFLELDSYHKIRKCIENVYGKDGIRMFNETDIDTFGYALTIFKDDED
IVAYLQNEYITKNGKRVSNLANKVYDKSLIDELLNLSFSKFAHLSMKAIRNILPYMEQGEIYS
KACELAGYNFTGPKKKEKALLLPVIPNIANPVVMRALTQSRKVVNAIIKKYGSPVSIHIELAR
DLSHSFDERKKIQKDQTENRKKNETAIKQLIEYELTKNPTGLDIVKFKLWSEQQGRCMYSLKP
IELERLLEPGYVEVDHILPYSRSLDDSYANKVLVLTKENREKGNHTPVEYLGLGSERWKKFEK
FVLANKQFSKKKKQNLLRLRYEETEEKEFKERNLNDTRYISKFFANFIKEHLKFADGDGGQKV
YTINGKITAHLRSRWDFNKNREESDLHHAVDAVIVACATQGMIKKITEFYKAREQNKESAKKK
EPIFPQPWPHFADELKARLSKFPQESIEAFALGNYDRKKLESLRPVFVSRMPKRSVTGAAHQE
TLRRCVGIDEQSGKIQTAVKTKLSDIKLDKDGHFPMYQKESDPRTYEAIRQRLLEHNNDPKKA
FQEPLYKPKKNGEPGPVIRTVKIIDTKNKVVHLDGSKTVAYNSNIVRTDVFEKDGKYYCVPVY
TMDIMKGTLPNKAIEANKPYSEWKEMTEEYTFQFSLFPNDLVRIVLPREKTIKTSTNEEIIIK
DIFAYYKTIDSATGGLELISHDRNFSLRGVGSKTLKRFEKYQVDVLGNIHKVKGEKRVGLAAP
TNQKKGKTVDSLQSVSDPKKKRKV

LrhCas9/NLS DNA sequence (SEQ ID NO:3)

ATGACCAAGCTGAACCAGCCTTACGGCATCGGCCTGGACATCGGCAGCAATAGCATCGGCTTT
GCCGTGGTGGACGCCAACAGCCATCTGCTGAGACTGAAGGGCGAGACAGCCATCGGCGCCAGA
CTGTTTAGAGAGGGACAGAGCGCCGCTGACAGACGGGGAAGCAGAACCACAAGAAGGCGGCTG
TCCAGAACCAGATGGCGGCTGAGCTTCCTGCGGGATTTCTTCGCCCCTCACATCACCAAGATC
GACCCCGACTTCTTTCTGCGGCAAAAATACTCCGAGATCAGCCCCAAGGACAAGGACAGGTTT
AAGTACGAGAAGCGGCTGTTCAACGACCGGACCGACGCCGAGTTCTACGAGGACTACCCCAGC
ATGTACCACCTGAGACTGCACCTGATGACCCACACACACAAGGCCGATCCTCGGGAAATCTTC
CTGGCCATCCACCACATCCTGAAGTCCAGAGGCCACTTTCTGACACCCGGCGCTGCCAAGGAC
TTCAACACCGACAAAGTGGACCTTGAGGACATCTTCCCCGCTCTGACAGAGGCTTACGCCCAG
GTGTACCCCGATCTGGAACTGACCTTCGATCTGGCCAAGGCCGACGACTTCAAGGCCAAGCTG
CTGGACGAACAGGCCACACCTAGCGACACACAGAAAGCCCTGGTCAACCTGCTGCTGTCTAGC
GACGGCGAGAAAGAAATCGTGAAGAAGCGGAAGCAGGTCCTGACCGAGTTCGCCAAGGCCATC
ACCGGCCTGAAAACAAAGTTCAATCTGGCCCTGGGCACCGAGGTGGACGAAGCTGATGCTTCC

AACTGGCAGTTCAGCATGGGCCAGCTGGACGACAAGTGGTCCAACATCGAAACCAGCATGACC
GACCAGGGCACCGAAATCTTCGAGCAGATCCAAGAGCTGTACCGGGCCAGACTGCTGAACGGA
ATTGTGCCTGCCGGCATGAGCCTGTCTCAGGCCAAAGTGGCCGATTACGGCCAGCACAAAGAG
GACCTGGAACTGTTCAAGACCTACCTGAAGAAGCTGAACGACCACGAGCTGGCCAAGACCATC
AGGGGCCTGTACGATCGGTACATCAACGGCGACGACGCCAAGCCTTTCCTGCGCGAGGATTTT
GTGAAGGCCCTGACCAAAGAAGTGACAGCTCACCCCAACGAGGTGTCCGAACAGCTGCTGAAC
AGGATGGGCCAAGCCAACTTCATGCTGAAGCAGCGGACCAAGGCCAACGGCGCCATTCCTATT
CAGCTGCAGCAGAGAGAGCTGGACCAGATCATTGCCAACCAGAGCAAGTACTACGACTGGCTG
GCCGCTCCTAATCCTGTGGAAGCCCACAGATGGAAGATGCCCTACCAGCTGGATGAGCTGCTC
AACTTTCACATCCCCTACTACGTGGGCCCTCTGATCACCCCTAAACAGCAGGCCGAGAGCGGC
GAGAATGTGTTCGCTTGGATGGTCCGAAAGGACCCCAGCGGCAACATCACCCCTTACAACTTC
GACGAGAAGGTGGACAGAGAGGCCAGCGCCAACACCTTCATCCAGAGAATGAAGACCACCGAC
ACATACCTGATCGGCGAGGACGTGCTGCCTAAGCAGAGCCTGCTGTACCAGAAATACGAGGTG
CTGAACGAGCTGAACAACGTGCGGATCAACAACGAGTGCCTGGGCACAGACCAGAAGCAGAGA
CTGATCAGAGAGGTGTTCGAGCGGCACAGCAGCGTGACCATCAAACAGGTGGCCGACAATCTG
GTGGCCCACGGCGATTTTGCCAGACGGCCTGAGATTAGAGGACTGGCCGATGAGAAGCGGTTC
CTGAGCAGCCTGAGCACCTACCACCAGCTGAAAGAGATCCTGCACGAGGCCATCGACGACCCC
ACCAAACTGCTGGATATCGAGAACATCATCACCTGGTCCACCGTGTTCGAGGACCACACCATC
TTCGAGACAAAGCTGGCCGAGATCGAGTGGCTGGACCCCAAGAAGATCAACGAGCTGTCTGGC
ATCAGATACAGAGGCTGGGGCCAGTTCTCCCGGAAGCTGCTCGATGGACTGAAGCTTGGCAAT
GGCCACACCGTGATTCAAGAACTGATGCTGAGCAACCACAACCTGATGCAGATCCTGGCCGAC
GAGACACTGAAAGAAACCATGACAGAGCTGAATCAGGACAAGCTGAAAACCGACGACATCGAG
GATGTGATCAACGACGCCTACACAAGCCCCAGCAACAAAAAGGCCCTCAGACAGGTGCTGAGA
GTGGTCGAGGATATCAAGCACGCCGCCAACGGACAGGACCCTAGCTGGCTGTTTATCGAAACC
GCCGATGGAACAGGCACCGCCGGCAAGAGAACACAGAGCCGGCAGAAACAGATCCAGACCGTG
TACGCCAACGCCGCTCAAGAGCTGATCGATTCTGCCGTGCGGGGCGAGCTGGAAGATAAGATT
GCTGACAAGGCCAGCTTCACCGACCGGCTGGTGCTGTACTTTATGCAAGGCGGCAGAGACATC
TACACAGGCGCCCCTCTGAACATCGACCAGCTGAGCCACTACGATATCGACCACATTCTGCCC
CAGAGCCTGATCAAGGACGACAGCCTGGACAACCGGGTGCTCGTGAACGCCACCATCAACCGC
GAGAAGAACAATGTGTTTGCCAGCACACTGTTCGCCGGAAAGATGAAGGCCACCTGGCGGAAA
TGGCACGAAGCCGGACTGATCTCTGGCAGAAAGCTGCGGAATCTGATGCTGCGGCCCGACGAG
ATCGACAAGTTTGCCAAGGGCTTCGTGGCCCGGCAGCTGGTTGAGACAAGACAGATCATCAAG
CTGACAGAGCAGATTGCCGCCGCTCAGTACCCCAACACCAAGATTATTGCCGTGAAGGCCGGA
CTGTCCCATCAGCTGAGAGAGGAACTGGACTTCCCCAAGAACCGGGACGTGAACCACTACCAC
CACGCCTTCGATGCCTTTCTGGCCGCTAGAATCGGCACCTACCTGCTGAAGAGATACCCCAAG
CTGGCCCCATTCTTCACCTACGGCGAGTTTGCTAAGGTGGACGTCAAGAAGTTCCGCGAGTTC
AACTTCATCGGAGCCCTGACACACGCCAAGAAGAACATTATCGCCAAGGACACCGGCGAGATC
GTGTGGGACAAAGAGCGGGACATCAGAGAACTGGACCGCATCTACAACTTCAAGCGGATGCTG
ATCACACACGAGGTGTACTTCGAGACTGCCGACCTGTTCAAGCAGACCATCTACGCCGCTAAG
GACAGCAAAGAGAGAGGCGGCAGCAAGCAGCTGATCCCTAAGAAGCAGGGCTACCCCACTCAG
GTGTACGGCGGCTACACACAAGAGAGCGGCTCTTACAACGCCCTCGTCAGAGTGGCCGAGGCC
GATACAACAGCCTACCAAGTGATCAAGATCAGCGCCCAGAACGCCAGCAAGATCGCCTCCGCC
AACCTGAAAAGCCGCGAGAAAGGCAAACAGCTCCTGAATGAGATCGTCGTGAAGCAGCTGGCT
AAGCGGCGGAAGAACTGGAAGCCTAGCGCCAATAGCTTCAAGATCGTGATCCCCAGATTCGGC
ATGGGCACCCTGTTCCAGAACGCTAAGTACGGCCTGTTCATGGTCAACAGCGACACCTACTAC
CGGAACTACCAAGAACTCTGGCTGAGCCGGGAAAACCAGAAACTGCTGAAAAAGCTGTTCTCC

ATCAAATACGAGAAAACCCAGATGAACCACGACGCCCTGCAGGTCTACAAGGCCATTATCGAC
CAGGTGGAAAAGTTCTTCAAGCTGTACGACATCAACCAGTTCCGCGCCAAGCTGAGCGACGCC
ATCGAGAGATTTGAGAAGCTGCCCATCAATACCGACGGCAACAAGATCGGCAAGACCGAGACT
CTGAGACAGATCCTGATCGGACTGCAGGCCAATGGCACCCGGTCCAACGTGAAGAACCTGGGC
ATCAAGACCGATCTGGGCCTGCTGCAAGTCGGCAGCGGAATCAAGCTGGACAAGGATACCCAG
ATCGTGTATCAGAGCCCCTCCGGCCTGTTTAAGCGGAGAATCCCACTGGCTGACCTGCCCAAG
AAGAAGAGGAAGGTG

LrhCas9/NLS protein sequence (SEQ ID NO:4)

MTKLNQPYGIGLDIGSNSIGFAVVDANSHLLRLKGETAIGARLFREGQSAADRRGSRTTRRRL
SRTRWRLSFLRDFFAPHITKIDPDFFLRQKYSEISPKDKDRFKYEKRLFNDRTDAEFYEDYPS
MYHLRLHLMTHTHKADPREIFLAIHHILKSRGHFLTPGAAKDFNTDKVDLEDIFPALTEAYAQ
VYPDLELTFDLAKADDFKAKLLDEQATPSDTQKALVNLLLSSDGEKEIVKKRKQVLTEFAKAI
TGLKTKFNLALGTEVDEADASNWQFSMGQLDDKWSNIETSMTDQGTEIFEQIQELYRARLLNG
IVPAGMSLSQAKVADYGQHKEDLELFKTYLKKLNDHELAKTIRGLYDRYINGDDAKPFLREDF
VKALTKEVTAHPNEVSEQLLNRMGQANFMLKQRTKANGAIPIQLQQRELDQIIANQSKYYDWL
AAPNPVEAHRWKMPYQLDELLNFHIPYYVGPLITPKQQAESGENVFAWMVRKDPSGNITPYNF
DEKVDREASANTFIQRMKTTDTYLIGEDVLPKQSLLYQKYEVLNELNNVRINNECLGTDQKQR
LIREVFERHSSVTIKQVADNLVAHGDFARRPEIRGLADEKRFLSSLSTYHQLKEILHEAIDDP
TKLLDIENIITWSTVFEDHTIFETKLAEIEWLDPKKINELSGIRYRGWGQFSRKLLDGLKLGN
GHTVIQELMLSNHNLMQILADETLKETMTELNQDKLKTDDIEDVINDAYTSPSNKKALRQVLR
VVEDIKHAANGQDPSWLFIETADGTGTAGKRTQSRQKQIQTVYANAAQELIDSAVRGELEDKI
ADKASFTDRLVLYFMQGGRDIYTGAPLNIDQLSHYDIDHILPQSLIKDDSLDNRVLVNATINR
EKNNVFASTLFAGKMKATWRKWHEAGLISGRKLRNLMLRPDEIDKFAKGFVARQLVETRQIIK
LTEQIAAAQYPNTKIIAVKAGLSHQLREELDFPKNRDVNHYHHAFDAFLAARIGTYLLKRYPK
LAPFFTYGEFAKVDVKKFREFNFIGALTHAKKNIIAKDTGEIVWDKERDIRELDRIYNFKRML
ITHEVYFETADLFKQTIYAAKDSKERGGSKQLIPKKQGYPTQVYGGYTQESGSYNALVRVAEA
DTTAYQVIKISAQNASKIASANLKSREKGKQLLNEIVVKQLAKRRKNWKPSANSFKIVIPRFG
MGTLFQNAKYGLFMVNSDTYYRNYQELWLSRENQKLLKKLFSIKYEKTQMNHDALQVYKAIID
QVEKFFKLYDINQFRAKLSDAIERFEKLPINTDGNKIGKTETLRQILIGLQANGTRSNVKNLG
IKTDLGLLQVGSGIKLDKDTQIVYQSPSGLFKRRIPLADLPKKKRKV

PexCas9/NLS DNA sequence (SEQ ID NO:5)

ATGGGCAAGACCCACATCATCGGCGTTGGCCTGGATCTCGGCGGCACATACACAGGCACCTTC
ATCACCAGCCATCCTAGCGACGAAGCCGAGCACAGAGATCACAGCAGCGCCTTCACCGTGGTC
AACAGCGAGAAGCTGAGCTTCAGCAGCAAGAGCAGAACAGCCGTGCGGCACAGAGTGCGGAGC
TACAAGGGCTTCGACCTGCGTAGAAGGCTGCTGCTTCTGGTGGCCGAGTATCAGCTGCTGCAG
AAGAAGCAGACACTGGCCCCTGAGGAAAGAGAGAACCTGAGAATCGCCCTGAGCGGCTACCTG
AAGAGAAGAGGCTACGCCAGAACCGAGGCCGAGACAGATACAAGCGTGCTGGAATCTCTGGAC
CCCAGCGTGTTCAGCAGCGCTCCCAGCTTCACCAATTTCTTCAACGACAGCGAGCCCCTGAAC
ATCCAGTGGGAAGCCATTGCCAACTCTCCCGAGACAACAAAGGCCCTGAACAAAGAGCTGAGC
GGCCAGAAAGAGGCCGACTTCAAGAAGTACATCAAGACCAGCTTTCCCGAGTACAGCGCCAAA
GAGATTCTGGCCAACTACGTGGAAGGCAGACGGGCCATTCTGGACGCCAGCAAGTATATCGCC
AACCTGCAGAGCCTGGGCCACAAGCACAGAAGCAAGTACCTGAGCGACATTCTGCAGGACATG
AAGCGGGACAGCCGGATCACCAGACTGAGCGAAGCCTTTGGCAGCACCGACAACCTGTGGCGG
ATCATCGGCAACATCAGCAATCTGCAAGAACGGGCCGTGCGGTGGTACTTCAACGATGCCAAG

TTCGAGCAGGGCCAAGAGCAGCTGGATGCCGTGAAGCTGAAGAATGTGCTCGTGCGGGCCCTG
AAGTATCTGCGGAGTGACGACAAAGAGTGGAGCGCCTCTCAGAAGCAGATCATCCAGTCTCTG
GAACAGAGCGGCGACGTGCTGGATGTGCTGGCTGGACTCGACCCCGACAGAACAATCCCTCCA
TACGAGGACCAGAACAACAGACGGCCTCCTGAGGATCAGACCCTGTATCTGAACCCCAAGGCT
CTGAGCAGCGAGTACGGCGAGAAGTGGAAGTCTTGGGCCAACAAGTTTGCCGGCGCTTACCCT
CTGCTGACCGAGGATCTGACCGAGATCCTGAAGAACACCGACAGAAAGTCCCGGATCAAGATC
AGATCCGATGTGCTGCCCGACAGCGACTACAGACTGGCCTACATCCTGCAGAGAGCCTTCGAT
CGGTCTATCGCCCTGGACGAGTGCAGCATCAGAAGAACCGCCGAGGACTTCGAGAACGGCGTG
GTCATCAAGAACGAGAAACTGGAAGATGTGCTGAGCGGACACCAGCTGGAAGAGTTTCTGGAA
TTTGCCAATCGGTACTACCAAGAGACAGCCAAGGCCAAGAACGGCCTGTGGTTCCCAGAGAAC
GCCCTGCTGGAAAGAGCCGATCTGCACCCTCCTATGAAGAACAAGATTCTGAACGTGATCGTC
GGACAGGCCCTGGGAGTGTCTCCTGCTGAGGGCACCGATTTCATCGAGGAAATTTGGAACAGC
AAAGTGAAAGGCCGGTCCACCGTGCGGAGCATCTGTAACGCCATCGAGAATGAGAGAAAGACC
TACGGACCCTACTTCAGCGAGGACTACAAGTTCGTGAAAACGGCCCTGAAAGAGGGCAAAACC
GAGAAAGAGCTGTCCAAGAAATTCGCCGCCGTGATCAAGGTGCTGAAGATGGTGTCTGAGGTG
GTGCCCTTTATCGGAAAAGAGCTGCGGCTGTCTGACGAGGCCCAGAGCAAGTTCGACAATCTG
TACTCTCTGGCCCAGCTGTACAACCTGATCGAGACAGAGCGGAACGGCTTCAGCAAGGTGTCA
CTGGCTGCCCACCTGGAAAATGCCTGGCGGATGACCATGACAGATGGATCCGCCCAGTGCTGT
AGACTGCCTGCCGATTGTGTGCGGCCCTTCGACGGCTTTATCCGGAAGGCCATCGACCGGAAC
TCTTGGGAAGTCGCCAAGCGGATTGCCGAGGAAGTGAAGAAGTCCGTCGACTTCACCAACGGC
ACCGTGAAGATCCCTGTGGCCATCGAGGCCAACAGCTTCAACTTTACCGCCAGCCTGACCGAC
CTGAAGTACATTCAGCTCAAAGAACAGAAGCTCAAGAAGAAGTTGGAGGACATCCAGCGGAAC
GAAGAGAATCAAGAGAAGCGGTGGCTGAGCAAAGAGGAACGGATCAGAGCCGACAGCCACGGC
ATCTGTGCCTATACTGGCAGACCCCTGGATGACGTGGGCGAGATCGATCACATCATCCCCAGA
AGCCTGACACTGAAGAAAAGCGAGAGCATCTACAACTCCGAAGTGAACCTGATCTTCGTGTCT
GCCCAGGGCAATCAAGAAAAGAAGAACAACATCTACCTGCTGAGCAACCTCGCCAAGAACTAC
CTGGCCGCCGTGTTTGGCACAAGCGACCTGAGCCAGATCACCAACGAGATCGAGAGCACCGTG
CTGCAGCTGAAAGCTGCTGGCAGACTGGGCTACTTCGATCTGCTGAGCGAAAAAGAGCGGGCC
TGCGCCAGACATGCCCTGTTTCTGAATAGCGACTCCGAGGCCAGACGCGCCGTGATTGATGTT
CTTGGCTCTCGGAGAAAGGCCAGCGTGAACGGAACCCAGGCTTGGTTTGTGCGGTCCATCTTC
TCCAAAGTGCGGCAGGCACTGGCCGCTTGGACACAAGAAACAGGCAACGAGCTGATCTTTGAC
GCCATCAGCGTGCCAGCCGCCGATAGCTCTGAGATGAGGAAGAGATTCGCCGAGTACCGGCCT
GAGTTCAGAAAGCCCAAAGTGCAGCCTGTGGCCTCTCACAGCATCGACGCCATGTGCATCTAT
CTGGCCGCCTGCAGCGACCCCTTCAAGACCAAGAGAATGGGCTCTCAGCTGGCCATCTACGAG
CCCATCAACTTCGATAACCTGTTCACCGGCAGCTGTCAAGTGATCCAGAACACCCCTCGGAAC
TTCTCCGACAAGACCAATATCGCTAACAGCCCCATCTTCAAAGAGACAATCTACGCCGAGCGG
TTCCTGGACATCATCGTGTCCAGAGGCGAGATTTTCATCGGCTACCCCAGCAACATGCCCTTC
GAGGAAAAGCCCAACCGGATCAGCATCGGCGGCAAGGACCCTTTCAGCATCCTGTCTGTGCTG
GGCGCCTACCTGGATAAGGCCCCTAGCAGCGAGAAAGAAAAGCTCACCATCTACCGGGTCGTC
AAGAACAAAGCCTTCGAGCTGTTCTCCAAGGTGGCCGGCAGCAAGTTTACCGCCGAAGAAGAT
AAGGCCGCCAAGATCCTGGAAGCCCTGCACTTCGTGACCGTGAAACAGGATGTGGCCGCCACC
GTGTCCGATCTGATCAAGAGCAAGAAAGAACTGAGCAAGGATAGCATCGAGAACCTGGCCAAG
CAGAAGGGCTGCCTGAAGAAGGTGGAATACTCCAGCAAAGAGTTCAAGTTCAAGGGCAGCCTG
ATCATCCCTGCCGCCGTGGAATGGGGAAAAGTGCTGTGGAACGTGTTCAAAGAAAACACGGCC
GAAGAACTGAAGGACGAGAACGCTCTGAGGAAGGCCCTGGAAGCTGCCTGGCCTAGCTCTTTC
GGCACCAGAAACCTGCACTCTAAGGCCAAGCGGGTGTTCAGCCTGCCTGTGGTGGCTACACAA

TCTGGCGCCGTGCGGATCAGACGCAAGACAGCCTTCGGCGACTTCGTGTACCAGAGCCAGGAC
ACAAACAACCTGTACAGCAGCTTCCCCGTGAAGAACGGCAAGCTGGATTGGAGCAGCCCTATC
ATTCACCCCGCTCTGCAGAACCGGAACCTGACCGCCTACGGCTACAGATTCGTGGACCACGAC
AGATCCATCAGCATGAGCGAGTTCAGAGAGGTGTACAACAAGGACGACCTGATGCGGATCGAG
CTGGCCCAGGGAACAAGCAGCAGACGCTACCTGAGAGTGGAAATGCCCGGCGAGAAATTCCTC
GCTTGGTTTGGCGAGAACAGCATCAGCCTGGGCTCCAGCTTCAAGTTCTCTGTGTCCGAGGTG
TTCGACAACAAAATCTACACCGAGAACGCCGAGTTTACCAAGTTCCTGCCTAAGCCTAGAGAG
GACAACAAGCACAACGGGACCATCTTTTTCGAACTCGTGGGCCCCAGAGTGATCTTCAACTAC
ATCGTTGGCGGAGCCGCCAGCAGCCTGAAAGAAATCTTTAGCGAGGCCGGCAAAGAGCGGAGC
CCCAAGAAGAAGAGGAAGGTG

PexCas9/NLS protein sequence (SEQ ID NO:6)

MGKTHIIGVGLDLGGTYTGTFITSHPSDEAEHRDHSSAFTVVNSEKLSFSSKSRTAVRHRVRS
YKGFDLRRRLLLLVAEYQLLQKKQTLAPEERENLRIALSGYLKRRGYARTEAETDTSVLESLD
PSVFSSAPSFTNFFNDSEPLNIQWEAIANSPETTKALNKELSGQKEADFKKYIKTSFPEYSAK
EILANYVEGRRAILDASKYIANLQSLGHKHRSKYLSDILQDMKRDSRITRLSEAFGSTDNLWR
IIGNISNLQERAVRWYFNDAKFEQGQEQLDAVKLKNVLVRALKYLRSDDKEWSASQKQIIQSL
EQSGDVLDVLAGLDPDRTIPPYEDQNNRRPPEDQTLYLNPKALSSEYGEKWKSWANKFAGAYP
LLTEDLTEILKNTDRKSRIKIRSDVLPDSDYRLAYILQRAFDRSIALDECSIRRTAEDFENGV
VIKNEKLEDVLSGHQLEEFLEFANRYYQETAKAKNGLWFPENALLERADLHPPMKNKILNVIV
GQALGVSPAEGTDFIEEIWNSKVKGRSTVRSICNAIENERKTYGPYFSEDYKFVKTALKEGKT
EKELSKKFAAVIKVLKMVSEVVPFIGKELRLSDEAQSKFDNLYSLAQLYNLIETERNGFSKVS
LAAHLENAWRMTMTDGSAQCCRLPADCVRPFDGFIRKAIDRNSWEVAKRIAEEVKKSVDFTNG
TVKIPVAIEANSFNFTASLTDLKYIQLKEQKLKKKLEDIQRNEENQEKRWLSKEERIRADSHG
ICAYTGRPLDDVGEIDHIIPRSLTLKKSESIYNSEVNLIFVSAQGNQEKKNNIYLLSNLAKNY
LAAVFGTSDLSQITNEIESTVLQLKAAGRLGYFDLLSEKERACARHALFLNSDSEARRAVIDV
LGSRRKASVNGTQAWFVRSIFSKVRQALAAWTQETGNELIFDAISVPAADSSEMRKRFAEYRP
EFRKPKVQPVASHSIDAMCIYLAACSDPFKTKRMGSQLAIYEPINFDNLFTGSCQVIQNTPRN
FSDKTNIANSPIFKETIYAERFLDIIVSRGEIFIGYPSNMPFEEKPNRISIGGKDPFSILSVL
GAYLDKAPSSEKEKLTIYRVVKNKAFELFSKVAGSKFTAEEDKAAKILEALHFVTVKQDVAAT
VSDLIKSKKELSKDSIENLAKQKGCLKKVEYSSKEFKFKGSLIIPAAVEWGKVLWNVFKENTA
EELKDENALRKALEAAWPSSFGTRNLHSKAKRVFSLPVVATQSGAVRIRRKTAFGDFVYQSQD
TNNLYSSFPVKNGKLDWSSPIIHPALQNRNLTAYGYRFVDHDRSISMSEFREVYNKDDLMRIE
LAQGTSSRRYLRVEMPGEKFLAWFGENSISLGSSFKFSVSEVFDNKIYTENAEFTKFLPKPRE
DNKHNGTIFFELVGPRVIFNYIVGGAASSLKEIFSEAGKERSPKKKRKV

McaCas9/NLS DNA sequence (SEQ ID NO:7)

ATGGAAAAGAAGCGGAAAGTCACCCTGGGCTTCGACCTGGGAATCGCCTCTGTTGGATGGGCC
ATCGTGGACAGCGAGACAAACCAGGTGTACAAGCTGGGCAGCAGACTGTTCGACGCCCCTGAC
ACCAACCTGGAAAGAAGAACCCAGCGGGGCACCAGAAGGCTGCTGCGGAGAAGAAAGTACCGG
AACCAGAAATTCTACAACCTGGTCAAGCGGACCGAGGTGTTCGGCCTGTCTAGCAGAGAGGCC
ATCGAGAACAGATTCAGAGAGCTGAGCATCAAGTACCCCAACATCATCGAGCTGAAAACAAAG
GCCCTGAGCCAAGAAGTGTGCCCCGACGAGATTGCCTGGATTCTGCACGACTACCTGAAGAAC
CGGGGCTACTTCTACGACGAGAAAGAGACAAAAGAGGACTTCGACCAGCAGACCGTGGAATCC
ATGCCTAGCTACAAGCTGAACGAGTTCTACAAGAAGTACGGCTACTTCAAAGGCGCCCTGTCT
CAGCCTACCGAGAGCGAGATGAAGGACAACAAGGACCTGAAAGAGGCATTCTTCTTCGACTTC

TCCAACAAAGAGTGGCTGAAAGAGATCAACTACTTCTTCAACGTGCAGAAGAACATCCTGAGC
GAGACATTCATCGAAGAGTTCAAGAAGATTTTCAGCTTCACCCGGGACATCAGCAAAGGCCCA
GGCAGCGACAATATGCCCTCTCCTTACGGCATCTTCGGCGAGTTCGGCGACAATGGCCAAGGC
GGCAGATACGAGCACATCTGGGACAAGAACATCGGCAAGTGCAGCATCTTCACCAACGAGCAG
AGAGCCCCTAAGTACCTGCCTAGCGCTCTGATCTTCAACTTCCTGAACGAGCTGGCCAACATC
AGACTGTACAGCACCGACAAGAAGAATATCCAGCCTCTGTGGAAGCTGAGCAGCATCGATAAG
CTGAATATCCTGCTGAACCTGTTCAACCTGCCTATCAGCGAGAAGAAGAAAAAGCTGACCAGC
ACCAACATCAACGACATCGTGAAGAAAGAGTCCATCAAGAGCATCATGCTGAGCGTCGAGGAC
ATCGACATGATCAAGGATGAGTGGGCCGGCAAAGAACCCAACGTGTACGGCGTTGGACTGAGC
GGCCTGAACATCGAGGAAAGCGCCAAAGAGAACAAGTTCAAGTTCCAAGACCTGAAGATCCTG
AACGTCCTGATCAATCTGCTGGACAACGTGGGCATCAAGTTCGAGTTCAAGGACCGCAGCGAC
ATCATCAAGAACCTGGAACTGCTGGATAACCTGTACCTGTTCCTGATCTACCAGAAAGAGAGC
AACAACAAAGACAGCTCCATCGACCTGTTTATCGCCAAGAACAAGTCCCTGAATATCGAGAAC
CTGAAGCTCAAGCTCAAAGAGTTCCTGCTCGGAGCCGGCAACGAGTTCGAGAACCACAACAGC
AAGACCCACAGCCTGTCCAAGAAGGCCATTGACGCCATCCTGCCTAAGCTGCTCGACAACAAC
GAAGGCTGGAATCTGGAAGCCATCAAGAATTACGACGAGGAAATCAAGAGCCAGATCGAGGAC
AACTCCAGCCTGATGGCCAAGCAGGATAAGAAGTACCTGAACGACAACTTCCTCAAGGATGCC
ATTCTGCCGCCAAACGTGAAAGTGACCTTCCAGCAGGCCATCCTCATCTTCAACAAGATCATC
CAGAAGTTCAGCAAGGATTTCGAGATCGACAAGGTCGTGATCGAACTGGCCAGAGAGATGACC
CAGGACCAAGAGAACGACGCCCTGAAGGGAATCGCTAAGGCCCAGAAGTCCAAGAAAAGCCTG
GTGGAAGAGAGACTCGAAGCCAACAACATCGACAAGAGCGTGTTCAACGATAAGTACGAGAAG
CTTATCTACAAGATTTTCCTGTGGATCAGCCAGGACTTTAAGGACCCCTACACCGGCGCCAAG
ATCAGCGCCAATGAGATCGTGGATAACAAGGTGGAAATCGACCACATCATCCCTTACAGCCTG
TGCTTCGACGACAGCAGCGCCAACAAAGTGCTGGTGCACAAGCAGAGCAATCAAGAGAAGTCT
AACAGCCTGCCGTACGAGTACATCAAGCAGGGCCACTCCGGCTGGAACTGGGACGAGTTCACC
AAATACGTGAAGCGGGTGTTCGTGAACAACGTGGACTCTATCCTGAGCAAGAAAGAGCGCCTG
AAGAAGTCCGAGAATCTGCTGACCACCAGCTACGACGGCTATGAGAAGCTGGGCTTCCTGGCC
AGAAACCTGAATGACACCAGATACGCCACCATCCTGTTCCGGGACCAGCTGAACAATTACGCC
GAGCACCACCTGATCGATAACAAGAAAATGTTCAAAGTGATCGCCATGAACGGGGCCGTGACC
AGCTTCATCCGGAAGAACATGAGCTACGACAACAAGCTGCGGCTGAAGGACAGAAGCGACTTC
AGCCACCACGCCTACGACGCCGCCATCATTGCCCTGTTCAGCAACAAGACCAAGACGCTGTAC
AACCTGATTGACCCCAGCCTGAACGGCATCATCAGCAAGAGAAGCGAAGGCTATTGGGTCATC
GAGGATCGGTACACAGGCGAGATCAAAGAGCTTAAGAAAGAGGATTGGACCTCTATCAAGAAC
AATGTGCAGGCCCGGAAGATCGCCAAAGAAATCGAGGAATATCTGATCGACCTGGACGATGAG
GTGTTCTTCAGCCGGAAAACTAAGCGCAAGACCAACCGGCAGCTGTACAATGAGACAATCTAC
GGAATCGCCGCCAAGACCGACGAGGACGGCATCACCAACTACTACAAGAAAGAAAAGTTCTCC
ATCCTGGACGACAAGGACATCTACCTGCGGCTGCTGAGAGAACGCGAGAAGTTCGTGATCAAC
CAGAGCAACCCCGAAGTGATCGACCAGATTATCGAGATCATCGAGAGCTACGGGAAAGAAAAC
AACATCCCCAGCCGCGACGAGGCCATCAATATCAAGTACACGAAGAACAAGATTAACTACAAC
CTCTACCTCAAGCAGTACATGCGGAGCCTGACCAAGAGCCTGGACCAGTTCAGCGAGGGCTTC
ATCAATCAGATGATCGCCAACAAGACGTTCGTGCTGTATAACCCCACCAAGAACACAACGCGG
AAGATCAAGTTCCTGCGGCTCGTGAACGATGTGAAGATCAACGATATTCGCAAGAATCAAGTG
ATCAACAAGTTTAACGGGAAGAACAACGAGCCCAAGGCCTTCTACGAGAATATCAACAGCCTG
GGCGCCATCGTGTTCAAGTCCTCCGCCAACAACTTCAAGACCCTGTCCATCAACACCCAGATC
GCCATCTTCGGAGACAAGAACTGGGATATCGAGGATTTCAAGACCTACAACATGGAAAAAATC
GAGAAGTACAAAGAGATATACGGCATCGACAAAACCTACAACTTCCACAGCTTTATCTTCCCC

GGCACAATCCTGCTCGATAAGCAGAACAAAGAGTTCTACTACATCAGCAGCATCCAGACCGTG
AACGACCAAATTGAGCTGAAGTTTCTGAACAAGATCGAGTTTAAGAACGACGACAACACCTCC
GGGGCCAACAAGCCTCCTCGGAGACTGAGATTCGGCATTAAGTCCATCATGAACAACTACGAG
CAGGTCGACATCAGCCCCTTCGGCATCAACAAGAAGATATTCGAGCCCAAGAAGAAGAGGAAG
GTG

McaCas9/NLS protein sequence (SEQ ID NO:8)

MEKKRKVTLGFDLGIASVGWAIVDSETNQVYKLGSRLFDAPDTNLERRTQRGTRRLLRRRKYR
NQKFYNLVKRTEVFGLSSREAIENRFRELSIKYPNIIELKTKALSQEVCPDEIAWILHDYLKN
RGYFYDEKETKEDFDQQTVESMPSYKLNEFYKKYGYFKGALSQPTESEMKDNKDLKEAFFFDF
SNKEWLKEINYFFNVQKNILSETFIEEFKKIFSFTRDISKGPGSDNMPSPYGIFGEFGDNGQG
GRYEHIWDKNIGKCSIFTNEQRAPKYLPSALIFNFLNELANIRLYSTDKKNIQPLWKLSSIDK
LNILLNLFNLPISEKKKKLTSTNINDIVKKESIKSIMLSVEDIDMIKDEWAGKEPNVYGVGLS
GLNIEESAKENKFKFQDLKILNVLINLLDNVGIKFEFKDRSDIIKNLELLDNLYLFLIYQKES
NNKDSSIDLFIAKNKSLNIENLKLKLKEFLLGAGNEFENHNSKTHSLSKKAIDAILPKLLDNN
EGWNLEAIKNYDEEIKSQIEDNSSLMAKQDKKYLNDNFLKDAILPPNVKVTFQQAILIFNKII
QKFSKDFEIDKVVIELAREMTQDQENDALKGIAKAQKSKKSLVEERLEANNIDKSVFNDKYEK
LIYKIFLWISQDFKDPYTGAKISANEIVDNKVEIDHIIPYSLCFDDSSANKVLVHKQSNQEKS
NSLPYEYIKQGHSGWNWDEFTKYVKRVFVNNVDSILSKKERLKKSENLLTTSYDGYEKLGFLA
RNLNDTRYATILFRDQLNNYAEHHLIDNKKMFKVIAMNGAVTSFIRKNMSYDNKLRLKDRSDF
SHHAYDAAIIALFSNKTKTLYNLIDPSLNGIISKRSEGYWVIEDRYTGEIKELKKEDWTSIKN
NVQARKIAKEIEEYLIDLDDEVFFSRKTKRKTNRQLYNETIYGIAAKTDEDGITNYYKKEKFS
ILDDKDIYLRLLREREKFVINQSNPEVIDQIIEIIESYGKENNIPSRDEAINIKYTKNKINYN
LYLKQYMRSLTKSLDQFSEGFINQMIANKTFVLYNPTKNTTRKIKFLRLVNDVKINDIRKNQV
INKFNGKNNEPKAFYENINSLGAIVFKSSANNFKTLSINTQIAIFGDKNWDIEDFKTYNMEKI
EKYKEIYGIDKTYNFHSFIFPGTILLDKQNKEFYYISSIQTVNDQIELKFLNKIEFKNDDNTS
GANKPPRRLRFGIKSIMNNYEQVDISPFGINKKIFEPKKKRKVPKKKRKV

MgaCas9/NLS DNA sequence (SEQ ID NO:9)

ATGAACAACAGCATCAAGAGCAAGCCCGAAGTGACCATCGGCCTGGATCTCGGCGTTGGCTCT
GTTGGATGGGCCATCGTGGACAACGAGACAAACATCATCCACCACCTGGGCAGCAGACTGTTC
AGCCAGGCCAAGACAGCTGAGGACAGGCGGTCTTTCAGAGGCGTGCGGAGACTGATCCGGCGG
AGAAAGTACAAGCTGAAGAGATTCGTGAACCTGATCTGGAAGTACAACAGCTACTTCGGCTTC
AAGAACAAAGAGGACATCCTGAACAACTACCAAGAGCAGCAGAAACTGCACAACACCGTGCTG
AACCTGAAGCTCGAAGCCCTGAACGCCAAGATCGACCCCAAGGCTCTGAGCTGGATTCTGCAC
GACTACCTGAAGAACCGGGGCCACTTCTACGAGGACAACCGGGACTTCAACGTGTACCCCACA
GAGGAACTGGCCAACTACTTCGACGAGTTCGGCTACTACAAGGGCATCATCGACAGCAAGAAC
GACGACGATGATAAGCTGGAAGAGGGCCTGACCAAGTACAAGTTCAGCAACCAGCACTGGCTG
GAAGAAGTGAAGAAGGTCCTGAGCAACCAGACCGGCCTGCCTGAGAAGTTCAAAGAGGAATAC
GAGAGCCTGTTCAGCTACGTGCGGAACTACTCTGAAGGCCCTGGCAGCATCAACAGCGTGTCC
CCATACGGCATCTATCACCTGGACGAGAAAGAGGGCAAAGTCGTCCAGAAGTATAACAACATC
TGGGACAAGACCATCGGGAAGTGCAGCATCTTCCCCGACGAGTACAGAGCCCCTAAGAACAGC
CCTATCGCCATGATCTTCAACGAGATCAACGAGCTGAGCACCATCCGGTCCTACAGCATCTAC
CTGACCGGCTGGTTCATCAATCAAGAGTTCAAGAAGGCCTACCTGAACAAGCTGCTGGACCTG
CTGATCAAGACCAACAGCGAGAAGCCCATCGACGCCCGGCAGTTTAAGAAGCTGCGGGAAGAG
ACAATCGCCGAGAGCATCGGCAAAGAAACCCTGAAGGACGTGGAAAGCGAGGAAAAGCTGGAA

AAGGACGACCACAAGTGGAAGCTGAAGGGCCTGAAGCTGAACACCAACGGCAAGATCCAGTAC
AACGACCTGTCTAGCCTGGCCAAGTTCGTGCACAAACTGAAGCAGCACCTGAAACTGGACTTT
CTGCTGGAAGATCAGTACACCCCTCTGGACAAGATCAACTTCCTGCAGAGCCTGTACGTGTAC
CTGGGCAAGCACCTGAGATACAGCAACAGAGTGGACAGCGCCAACCTGAAAGAGTTCAGCGAC
AGCTCCCGGCTGTTCGAGAGAGTGCTGCAAGAGCAGAAGGACGGCCTGTTCAAGCTGTTTGAG
CAGACCGACAAGGACGACGAGAAGATCCTGACACAGACCCACAGCCTGTCCACCAAGGCTATG
CTGCTGGCCATCACCAGAATGACCAACCTGGACAATGACGAGGATAACCAGAAGAACAACGAC
AAAGGCTGGAACTTCGAGGCCATCAAGAACTTCGACCAGAAGTTCATCGACATCACCAAGACG
AACAACAACCTGAGCCTGAAGCAGGACAAGCGCTACCTGGATGACCAGTTCATCAACGACGCC
ATTCTGAGCCCTGGCGTGAAGAGAATCCTGCGCGAGGCCACCAAGGTGTTCAACGCCATCCTC
AAGCAGTTCTCCGAAGAGTACGACGTGACCAAGGTGGTCATCGAGCTGGCCAGAGAGCTGAGC
GAAGAGAAAGAACTGGAAAACACCAAGAACTACAAGAAGCTTATCAAGAAGAACGGCGATAAG
ATCAGCGAGGGACTGAAAGCCCTGGGGATCGCCGAGGATAAGATCGAAGAGATCCTGAAGTCT
CCCACCAAGTCCTACAAAGTGCTGCTGTGGCTGCAGCAGGACCACATCGATCCCTACAGCCAG
AAAGAGATCGCCTTCGACGATATCCTGACCAAAACCGAAAAGACCGAGATCGACCACATCATT
CCTTACTCCATCAGCTTCGACGACAGCAGCAGCAACAAACTGCTGGTGCTGGCCGAGTCCAAT
CAGGCCAAGTCCAACCAGACACCTTACGAGTTTATCAACTCCGGCAAGGCCGAGATCACCTGG
GAAGTGTACGAGGCCTACTGCCACAAGTTCAAAAACGGCGACTCCAGCCTGCTGGACAGCACC
CAGAGAAGCAAGAAATTCGCCAAGATGATGAAGACCGACACCAGCTCTAAGTACGACATCGGC
TTTCTGGCCCGGAACCTGAACGACACCAGATACGCCACCATCGTGTTCCGGGACGCTCTGAAG
GACTACGCCAACAACCACCTGGTGGAAGATAAGCCCATGTTCAAGGTCGTGTGCATCAACGGC
GGCGTGACCAGCTTCCTGCGGAAGAACTTTGACCCCAAGTCTTGGTACGCCAAGAAGGACAGA
GACAAGAACATTCACCACGCCGTGGACGCCAGCATCATCTCCATCTTCAGCAACGAGACTAAG
ACCCTGTTCAACCAGCTGACAAAGTTCGCCGACTACAAGCTGTTCAAGAATACCGACGGCTCT
TGGAAGAAGATCGATCCTAAGACAGGCGTGGTGTCAGAAGTGACCGACGAGAATTGGAAGCAG
ATCCGCGTGCGCAACCAGGTGTCCGAGATCGCCAAAGTGATCGACAAGTACATCCAGGACAGC
AACATCGAGCGGAAGGCCAGATACAGCCGGAAGATCGAGAACAAGACCAATATCAGCCTGTTT
AACGACACCGTGTACTCCGCCAAGAAAGTGGGCTACGAGGATCAGATCAAGCGCAAGAACCTG
AAAACCCTGGACATCCACGAGAGCGCCGAGGAAAACAAGAACAGCAAAGTGAAAAAGCAGTTC
GTGTACCGGAAGCTCGTGAACGTGTCCCTGCTGAACAATGACAAGCTGGCCGACCTGTTCGCC
GAGAAAGAAGATATTCTGATGTACCGGGCCAATCCGTGGGTCATCAACCTGGCCGAGCAGATT
TTCAACGAGTACACCGAGAACAAAAAGATCAAGAGCCAGAACGTGTTCGAGAAGTACATGCTG
GATCTGACCAAAGAGTTCCCCGAGAAGTTTAGCGAGGCCTTCGTGAAGTCCATGATCAGAAAC
AAGACCGCCATCATCTACAACGTCGAGAAGGATGTGGTGCACCGGATCAAGCGGCTGAAGATT
CTGAGCAGCGAGCTGAAAGAAAACAAGTGGTCCAACGTGATCATCCGCTCCAAGAACGAGAGC
GGCACCAAGCTGAGCTACCAGGACACCATCAACTCTATCGCCCTGATGATCATGCGGAGCATC
GACCCAACCGCCAAAAAACAGTACATCAGGGTGCCCCTGAACACCCTGAATCTGCACCTGGGC
GACCAGGACTTCGACCTGCACAATATCGACGCCTATCTGAAGAAGCCTAAGTTCGTCAAGTAC
CTGAAGGCCAATGAGATCGGCGACGAGTATAAGCCTTGGCGCGTGCTGACAAGCGGCACACTG
CTGATCCACAAGAAAGACAAGAAACTCATGTACATCAGCAGCTTCCAGAACCTCAACGACCTC
ATCGAGATCAAGAATCTGATCGAGACAGAGTACAAAGAAAACGTGGACTCAGACCCCAAGAAG
AAGAAAAAGGCCAGCCAGATCCTGAGAAGCCTGAGCGTGATCCTGAACGATTACATCCTGCTG
GATGCCAAGTATAACTTCGACATCCTGGGCCTGTCTAAGAACAAGATTGACGAGATCCTCAAC
AGCAAGCTGGACCTCGACAAGATTGCCAAGCCCAAGAAGAAGAGGAAGGTG

MgaCas9/NLS protein sequence (SEQ ID NO:10)

MNNSIKSKPEVTIGLDLGVGSVGWAIVDNETNIIHHLGSRLFSQAKTAEDRRSFRGVRRLIRR
RKYKLKRFVNLIWKYNSYFGFKNKEDILNNYQEQQKLHNTVLNLKLEALNAKIDPKALSWILH
DYLKNRGHFYEDNRDFNVYPTEELANYFDEFGYYKGIIDSKNDDDDKLEEGLTKYKFSNQHWL
EEVKKVLSNQTGLPEKFKEEYESLFSYVRNYSEGPGSINSVSPYGIYHLDEKEGKVVQKYNNI
WDKTIGKCSIFPDEYRAPKNSPIAMIFNEINELSTIRSYSIYLTGWFINQEFKKAYLNKLLDL
LIKTNSEKPIDARQFKKLREETIAESIGKETLKDVESEEKLEKDDHKWKLKGLKLNTNGKIQY
NDLSSLAKFVHKLKQHLKLDFLLEDQYTPLDKINFLQSLYVYLGKHLRYSNRVDSANLKEFSD
SSRLFERVLQEQKDGLFKLFEQTDKDDEKILTQTHSLSTKAMLLAITRMTNLDNDEDNQKNND
KGWNFEAIKNFDQKFIDITKTNNNLSLKQDKRYLDDQFINDAILSPGVKRILREATKVFNAIL
KQFSEEYDVTKVVIELARELSEEKELENTKNYKKLIKKNGDKISEGLKALGIAEDKIEEILKS
PTKSYKVLLWLQQDHIDPYSQKEIAFDDILTKTEKTEIDHIIPYSISFDDSSSNKLLVLAESN
QAKSNQTPYEFINSGKAEITWEVYEAYCHKFKNGDSSLLDSTQRSKKFAKMMKTDTSSKYDIG
FLARNLNDTRYATIVFRDALKDYANNHLVEDKPMFKVVCINGGVTSFLRKNFDPKSWYAKKDR
DKNIHHAVDASIISIFSNETKTLFNQLTKFADYKLFKNTDGSWKKIDPKTGVVSEVTDENWKQ
IRVRNQVSEIAKVIDKYIQDSNIERKARYSRKIENKTNISLFNDTVYSAKKVGYEDQIKRKNL
KTLDIHESAEENKNSKVKKQFVYRKLVNVSLLNNDKLADLFAEKEDILMYRANPWVINLAEQI
FNEYTENKKIKSQNVFEKYMLDLTKEFPEKFSEAFVKSMIRNKTAIIYNVEKDVVHRIKRLKI
LSSELKENKWSNVIIRSKNESGTKLSYQDTINSIALMIMRSIDPTAKKQYIRVPLNTLNLHLG
DQDFDLHNIDAYLKKPKFVKYLKANEIGDEYKPWRVLTSGTLLIHKKDKKLMYISSFQNLNDL
IEIKNLIETEYKENVDSDPKKKKKASQILRSLSVILNDYILLDAKYNFDILGLSKNKIDEILN
SKLDLDKIAKPKKKRKV

AglCas9/NLS DNA sequence (SEQ ID NO:11)

ATGCAGAACATCACCTTCAGCTTCGACGTGGGCTACGCCTCTATCGGATGGGCTGTTGTTCAG
GCCCCTGCTCAGCCAGAGCAGGACCCTGGAATAGTGGCCTGTGGCACCGTGCTGTTCCCTAGC
GACGATTGCCAGGCCTTCCAGCGGAGAAACTACCGGCGGCTGCGGAGGAACATCCGGTCCAGA
AGAGTGCGGATCGAGCGGATCGGAAAGCTGCTGGTTCAGGCCGGAATCCTGACACCTGAGGAA
AAGGCCACACCTGGACACCCCGCTCCATTCTTTTTGGCAGCCCAGGCTTGGCAGGGCATCAGA
CAACTGTCTCCACTGGAAGTGTGGCACATCCTGCGTTGGTACGCCCACAACAGAGGCTACGAC
AACAATGCCGCCTGGGCCACCGTGTCCACCAAAGAGGATACCGAGAAAGTCAACAACGCCCGG
CACCTGATGCAGAAGTTTGGCGCCGAGACAATGTGCGCCACACTGTGCCATGCTATGGAACTG
GACATGGACGTGCCCGATGCCGCCATGACAGTGTCTACACCAGCCTACAGAACCCTGAACAGC
GCCTTTCCTAGAGATGTGGTGCAGAGAGAGGTGCTGGACATCCTGAGACACAGCGCCAGCCAC
ATCAAAGAGCTGACCCCTGAGATCATCCGGCTGATCGCCCAGCAAGAGGATCTGAGCACAGAG
CAGAGAAGCGAGCTGGCCGCCAAGGGAATTAGACTGGCCAGAAGATACCGGGGCAGCCTGCTG
TTTGGACAGCTGCTGCCCAGATTCGACAACCGGATCATCAGACGGTGCCCCATCATCTGGGCC
CACACATTTGAGCAGGCCAAGACCAGCGGCATGAGCGAGAAAGAAGCTCAGGCCCTGGCTGAC
AAGGTGGCCAAAGTGCCTACAGCCGACTGTCCCGAGTTCTACGCCTACAGATTCGCCCGCATC
CTGAACAATCTGAGAGCCAACGGACTGCCCCTGCCTGTGGAAGTTCGCTGTGAACTGATGCAG
GCCGCCAGAGCCGAGGGAAAACTGACAGCCGCCAAGATCAAGAAAGAAATCATGAGGCTGATG
GGCGACGTCGAGAGCAACATCGACCACTACTTCCATCTGCACCCCGACAGCGAGGAAGCCCTG
ATTCTCGATCCCGCTATGGAATGCCTGCACCGGACCGGACTGTACGATGCCCTCAGCTCTGTC
GTGCGAAGAGTGGCCCTGACCAGACTGCGGAGAGGCAAAATCTGTACCCCTGCCTACCTGCGG
GACATGATGCTGAGACACGGCGAGGATACCCAGGCTCTGGATCTGGCCATTGCCAAGCAGCAG
GGAAGAAAGGCCCCTCGGCCTAGAAAGAACGACACAGATGCCAGCGCCGACGCCAGCATTGCA
TGGCAAGATAAGCCCCTGGCTCCTAAGACAGCCTCTGGCAGAGCCCCTTATGCCAGACCAGTT

CTGAGACAGGCCGTGGACGAGATCATGAATGGCGAGGACCCTACCAGGCCAGCTCTGGATGAA
CAGCATCCCGACGGCGAGGACAAGCCTTCTCACGGCTGTCTGTATGGCCTGCTGGACCCTGCC
AGCAAAGAGAACGAGTACCTGAACAAGCTTCCCCTGGACGCCCTGACAAACAATCACCTCGTG
CGGCACCGGATGCTGATCCTGGATAGACTGACCCAGGACCTCGTCAGAGAGTTCGCTGACGGC
GATCCCAGCAGAGTGGAACGGTTCTGTATCGAAGTGGGCAGAGAGCTGAGCGCCTTCTCTGGC
ATGACCAGCAAGCAGATCCAGTCCGAGCTGAACGAGCGGATGAAGCACTTCAAGAGCGCCGTG
GCCTATCTGGCCAAACACGCCCCTGATATGGCCACATCTGCCGGCCTGATCCGGAAGTGCAGA
ATCGCTATGGACATGAACTGGCAGTGCCCTTTCACCGGCCAGACCTACATGCCCTACGACCTG
CCTAAGCTGGAACGCGAGCACATCGTGCCCTACGCCAACAGAAAGACAGATGCCCTGTCTGCC
CTGGTGCTGACATGGCTGGCCGTGAACAAGATGAAGGGCAAGAGAACCGCCTACCAGTTTATC
AAAGAGTGCGAGGGCCAGAGCGTGCCCGGCAGAAATCAGAATATCGTGTCCGTGAAGCAGTAC
GAGACATTCGTGGAAAAGCTGGACACCAAGGGCCACGCCGACGACGCCAAGAGAAAAAAGACC
CGGAAGAAACTGATGATGGTGGACAGACTGAGCAGCCAGGGAACAAACGGCGAGTCTGAGCTG
GATTTCACCGAGGGCATGATGACCCAGAGCAGCCACCTGATGAAGATCGCCGCTAGAGGCGTG
CGGAAGAACTTTCCTCACGCCACCGTGGACATGATCCCTGGCGCTATTACTGGCACTGTGCGC
AAGGCTTGGAAGGTGGCAGGATGCCTGGCCGGCATTTGTCCTGAAGCCGTCGATCCCGTGACA
CACAGAATCCAGGACAAAGAGACACTGCGGCGGCTGACCCATCTGCATCATGCACTGGATGCC
TGCGTGCTGGGACTGATCCCTCACCTGATTCCAGAGCACAGATCCGGCCTGCTGAGAAAAGCT
CTGGCCGCTAGAAGGCTGCCCGAGAATGTTCGGCAAGAGGTGGAAAGCGCCGTGTCCAAGCGG
TACTACACCATCACAAAAGAGAGCAAACTGGAACTGCGGGATCTGCCCACCACACTGAAGAAC
TCTATCGCCGCCAAGCTGAGCGAGGGCAGAGTGGTGCAACACATCCCTGCCGATATGAGCGGA
GCCAAGCTGGAAGAGACAATCTGGGGAATTGCCCCTGGCCAGCACATCGACGACAATAGCGAG
GTGGTCATCCGGCAGAAGTCCCTGAGCATCGGCAAGGACGGCAACAGAATCAGAACCAGAAAG
ACCGACAAGCAGGGCAACCCCATCACCGAGAAGGCCTCTAAGCTCGTGGGCATCAAGCCTACC
GGCACCAGCAAACTGCAGCCCATCAGAGGCGTGATCATCATCAAGGACAACTTCGCCATTGCT
CTGGACCCCGTGCCAACCATGATTCCCCACCACAACGTGTACAAGCGGCTGGAAGAACTGCGG
AAGCTGAACCACGGTAGACATGTGCGGCTGCTGAAAAAGGGCATGCTGATCAGGCTGAGCCAC
CAGAAGTCCGGCGACAAGAACGGCATGTGGAAAGTGCGGAGCATCCAGGACCAGGGCTCCTCT
GGCCTGAAAGTGAATCTGCAGAGGCCCTACTACGCCGGCAAGATCGAGGACACCAGAACCGAG
AATTGGAAGAACGTGTCCATCAAGGCCCTGCTGAGCCAAGGCATGGAAATCCTGCCAACCACC
TACTGCGGCACCACACCTCCCAAGAAGAAGAGGAAGGTG

AglCas9/NLS protein sequence (SEQ ID NO:12)

MQNITFSFDVGYASIGWAVVQAPAQPEQDPGIVACGTVLFPSDDCQAFQRRNYRRLRRNIRSR
RVRIERIGKLLVQAGILTPEEKATPGHPAPFFLAAQAWQGIRQLSPLEVWHILRWYAHNRGYD
NNAAWATVSTKEDTEKVNNARHLMQKFGAETMCATLCHAMELDMDVPDAAMTVSTPAYRTLNS
AFPRDVVQREVLDILRHSASHIKELTPEIIRLIAQQEDLSTEQRSELAAKGIRLARRYRGSLL
FGQLLPRFDNRIIRRCPIIWAHTFEQAKTSGMSEKEAQALADKVAKVPTADCPEFYAYRFARI
LNNLRANGLPLPVEVRCELMQAARAEGKLTAAKIKKEIMRLMGDVESNIDHYFHLHPDSEEAL
ILDPAMECLHRTGLYDALSSVVRRVALTRLRRGKICTPAYLRDMMLRHGEDTQALDLAIAKQQ
GRKAPRPRKNDTDASADASIAWQDKPLAPKTASGRAPYARPVLRQAVDEIMNGEDPTRPALDE
QHPDGEDKPSHGCLYGLLDPASKENEYLNKLPLDALTNNHLVRHRMLILDRLTQDLVREFADG
DPSRVERFCIEVGRELSAFSGMTSKQIQSELNERMKHFKSAVAYLAKHAPDMATSAGLIRKCR
IAMDMNWQCPFTGQTYMPYDLPKLEREHIVPYANRKTDALSALVLTWLAVNKMKGKRTAYQFI
KECEGQSVPGRNQNIVSVKQYETFVEKLDTKGHADDAKRKKTRKKLMMVDRLSSQGTNGESEL
DFTEGMMTQSSHLMKIAARGVRKNFPHATVDMIPGAITGTVRKAWKVAGCLAGICPEAVDPVT

HRIQDKETLRRLTHLHHALDACVLGLIPHLIPEHRSGLLRKALAARRLPENVRQEVESAVSKR
YYTITKESKLELRDLPTTLKNSIAAKLSEGRVVQHIPADMSGAKLEETIWGIAPGQHIDDNSE
VVIRQKSLSIGKDGNRIRTRKTDKQGNPITEKASKLVGIKPTGTSKLQPIRGVIIIKDNFAIA
LDPVPTMIPHHNVYKRLEELRKLNHGRHVRLLKKGMLIRLSHQKSGDKNGMWKVRSIQDQGSS
GLKVNLQRPYYAGKIEDTRTENWKNVSIKALLSQGMEILPTTYCGTTPPKKKRKV

AmuCas9/NLS DNA sequence (SEQ ID NO:13)

ATGAGCAGAAGCCTGACCTTCAGCTTCGACATCGGCTACGCCTCTATCGGCTGGGCCGTGATT
GCCTCTGCCAGCCACGATGATGCCGATCCTAGCGTGTGTGGCTGTGGCACCGTGCTGTTCCCC
AAGGATGATTGCCAGGCCTTCAAGCGGAGAGAGTACCGGCGGCTGCGGAGAAACATCCGGTCC
AGAAGAGTGCGGATCGAGCGGATTGGTAGACTGCTGGTGCAGGCCCAGATCATCACCCCTGAG
ATGAAGGAAACCAGCGGACACCCCGCTCCATTCTACCTGGCATCTGAGGCCCTGAAGGGCCAC
AGAACACTGGCCCCTATTGAACTGTGGCATGTGCTGCGTTGGTACGCCCACAACAGAGGCTAC
GACAACAACGCCAGCTGGTCCAACAGCCTGTCTGAGGATGGTGGCAACGGCGAGGATACCGAG
AGAGTGAAACACGCCCAGGACCTGATGGACAAGCACGGCACAGCTACAATGGCCGAGACAATC
TGCAGAGAGCTGAAGCTGGAAGAGGGCAAAGCCGACGCTCCTATGGAAGTGTCTACCCCTGCC
TACAAGAACCTGAACACCGCCTTTCCACGGCTGATCGTGGAAAAAGAAGTGCGGAGAATCCTG
GAACTGAGCGCCCCTCTGATCCCTGGACTGACAGCCGAGATCATCGAGCTGATCGCCCAGCAT
CACCCTCTGACCACTGAACAGAGAGGCGTGCTGCTCCAGCACGGCATTAAGCTGGCCAGAAGA
TACAGAGGCAGCCTGCTGTTCGGCCAGCTGATCCCTAGATTCGACAACAGGATCATCAGCAGA
TGCCCCGTGACATGGGCCCAAGTGTATGAGGCCGAGCTGAAGAAGGGCAACAGCGAGCAGTCT
GCCAGAGAGAGAGCCGAGAAGCTGAGCAAGGTGCCCACCGCCAATTGTCCCGAGTTCTACGAG
TACCGGATGGCCAGAATCCTGTGCAACATCAGAGCCGACGGCGAGCCTCTGAGCGCCGAGATT
AGACGCGAGCTGATGAACCAGGCCAGACAAGAGGGAAAGCTGACCAAGGCCAGCCTGGAAAAG
GCCATCTCTAGCCGGCTGGGCAAAGAAACCGAGACAAACGTGTCCAACTACTTCACACTGCAC
CCCGACAGCGAGGAAGCCCTGTATCTGAATCCTGCCGTGGAAGTGCTGCAGAGAAGCGGCATC
GGCCAGATTCTGAGCCCCAGCGTGTACAGAATCGCCGCCAACAGACTGCGGAGAGGCAAGAGC
GTGACCCCTAACTACCTGCTGAATCTGCTGAAGTCCAGAGGCGAGTCTGGCGAGGCCCTGGAA
AAAAAGATCGAGAAAGAGTCCAAGAAGAAAGAGGCCGACTACGCCGACACACCCCTGAAGCCT
AAGTACGCCACAGGCAGAGCCCCTTACGCCAGAACCGTGCTGAAGAAAGTGGTGGAAGAGATC
CTGGATGGCGAGGACCCTACCAGACCTGCTAGAGGCGAAGCTCACCCTGACGGCGAACTGAAA
GCCCACGATGGCTGCCTGTACTGCCTGCTGGATACCGACAGCAGCGTGAACCAGCACCAGAAA
GAGCGGAGACTGGACACCATGACCAACAACCACCTCGTGCGGCACCGGATGCTGATCCTGGAC
AGACTCCTGAAGGATCTGATCCAGGACTTCGCCGACGGCCAGAAGGACAGAATCAGCAGAGTG
TGCGTGGAAGTCGGCAAAGAGCTGACCACCTTCAGCGCTATGGACAGCAAGAAGATCCAGCGG
GAACTGACCCTGCGGCAGAAGTCTCATACCGACGCCGTGAACAGACTGAAGAGAAAGCTTCCA
GGCAAGGCCCTGAGCGCCAACCTGATCAGAAAGTGCAGAATCGCAATGGACATGAACTGGACA
TGCCCCTTCACCGGCGCCACATATGGCGATCACGAGCTGGAAAATCTGGAACTGGAACACATC
GTGCCCCACAGCTTCAGACAGAGCAATGCCCTGTCTAGCCTGGTGCTGACATGGCCTGGCGTG
AACAGGATGAAGGGACAGAGAACCGGCTACGACTTCGTGGAACAAGAGCAAGAGAACCCCGTG
CCTGACAAGCCCAACCTGCACATCTGCAGCCTGAACAACTATCGCGAGCTGGTGGAAAAGCTG
GACGACAAGAAGGGACACGAGGACGACAGACGGCGGAAGAAGAAAAGAAAGGCCCTGCTGATG
GTCCGAGGCCTGTCTCACAAACACCAGAGCCAGAACCACGAGGCCATGAAAGAAATCGGCATG
ACCGAGGGCATGATGACCCAGAGCAGCCACCTGATGAAGCTGGCCTGCAAGAGCATCAAGACC
AGCCTGCCTGACGCTCACATCGACATGATTCCAGGCGCCGTGACTGCCGAAGTTCGCAAAGCC
TGGGATGTGTTCGGCGTGTTCAAAGAACTGTGCCCCGAAGCCGCCGATCCTGACTCTGGCAAG

ATCCTGAAAGAGAACCTGCGGAGCCTGACTCATCTGCATCACGCCCTGGATGCCTGTGTGCTG
GGACTGATCCCCTACATCATCCCCGCTCACCACAATGGCCTGCTGAGAAGAGTCCTGGCCATG
CGCAGAATCCCCGAGAAACTGATCCCTCAAGTGCGGCCCGTGGCCAACCAGAGACACTACGTG
CTGAACGACGACGGCCGGATGATGCTGAGGGATCTGAGTGCCAGCCTGAAAGAAAACATCCGC
GAGCAGCTGATGGAACAGCGAGTGATCCAGCACGTGCCCGCTGATATGGGCGGAGCACTGCTC
AAAGAAACAATGCAGCGGGTGCTGAGCGTGGACGGCTCTGGCGAAGATGCTATGGTGTCCCTG
TCTAAGAAGAAGGACGGCAAGAAAGAGAAGAATCAAGTCAAGGCCTCCAAGCTCGTGGGAGTG
TTTCCTGAGGGCCCCAGCAAGCTGAAAGCTCTGAAGGCCGCCATCGAGATCGACGGCAATTAT
GGCGTGGCACTGGACCCCAAGCCTGTGGTCATCAGACACATCAAGGTGTTCAAGAGGATCATG
GCCCTCAAAGAGCAGAACGGCGGCAAGCCAGTGCGCATCCTGAAAAAGGGCATGCTGATTCAC
CTGACCAGCAGCAAGGACCCTAAGCACGCTGGCGTTTGGAGAATCGAGAGCATCCAGGACAGC
AAAGGCGGCGTGAAACTGGACCTGCAGAGGGCTCATTGCGCCGTGCCTAAGAACAAGACCCAC
GAGTGCAATTGGAGAGAGGTGGACCTGATCTCCCTGCTGAAAAAGTACCAGATGAAGCGCTAC
CCCACCAGCTACACCGGCACACCTAGACCCAAGAAGAAGAGGAAGGTG

AmuCas9/NLS protein sequence (SEQ ID NO:14)

MSRSLTFSFDIGYASIGWAVIASASHDDADPSVCGCGTVLFPKDDCQAFKRREYRRLRRNIRS
RRVRIERIGRLLVQAQIITPEMKETSGHPAPFYLASEALKGHRTLAPIELWHVLRWYAHNRGY
DNNASWSNSLSEDGGNGEDTERVKHAQDLMDKHGTATMAETICRELKLEEGKADAPMEVSTPA
YKNLNTAFPRLIVEKEVRRILELSAPLIPGLTAEIIELIAQHHPLTTEQRGVLLQHGIKLARR
YRGSLLFGQLIPRFDNRIISRCPVTWAQVYEAELKKGNSEQSARERAEKLSKVPTANCPEFYE
YRMARILCNIRADGEPLSAEIRRELMNQARQEGKLTKASLEKAISSRLGKETETNVSNYFTLH
PDSEEALYLNPAVEVLQRSGIGQILSPSVYRIAANRLRRGKSVTPNYLLNLLKSRGESGEALE
KKIEKESKKKEADYADTPLKPKYATGRAPYARTVLKKVVEEILDGEDPTRPARGEAHPDGELK
AHDGCLYCLLDTDSSVNQHQKERRLDTMTNNHLVRHRMLILDRLLKDLIQDFADGQKDRISRV
CVEVGKELTTFSAMDSKKIQRELTLRQKSHTDAVNRLKRKLPGKALSANLIRKCRIAMDMNWT
CPFTGATYGDHELENLELEHIVPHSFRQSNALSSLVLTWPGVNRMKGQRTGYDFVEQEQENPV
PDKPNLHICSLNNYRELVEKLDDKKGHEDDRRRKKKRKALLMVRGLSHKHQSQNHEAMKEIGM
TEGMMTQSSHLMKLACKSIKTSLPDAHIDMIPGAVTAEVRKAWDVFGVFKELCPEAADPDSGK
ILKENLRSLTHLHHALDACVLGLIPYIIPAHHNGLLRRVLAMRRIPEKLIPQVRPVANQRHYV
LNDDGRMMLRDLSASLKENIREQLMEQRVIQHVPADMGGALLKETMQRVLSVDGSGEDAMVSL
SKKKDGKKEKNQVKASKLVGVFPEGPSKLKALKAAIEIDGNYGVALDPKPVVIRHIKVFKRIM
ALKEQNGGKPVRILKKGMLIHLTSSKDPKHAGVWRIESIQDSKGGVKLDLQRAHCAVPKNKTH
ECNWREVDLISLLKKYQMKRYPTSYTGTPRPKKKRKV

SEQ ID NO:15. OkiCas9/NLS DNA

ATGGCCAGAGATTACAGCGTCGGCCTGGATATCGGCACCTCTTCTGTTGGATGGGCCGCCATC
GACAACAAGTACCACCTGATCCGGGCCAAGAGCAAGAACCTGATTGGCGTGCGGCTGTTCGAT
AGCGCCGTGACCGCCGAGAAGAGAAGAGGCTACAGAACCACCAGACGGCGGCTGAGCAGACGG
CATTGGAGACTGAGACTGCTGAACGACATCTTCGCCGGACCTCTGACCGATTTCGGCGACGAG
AATTTCCTGGCCAGACTGAAGTACAGCTGGGTTCACCCTCAAGACCAGAGCAATCAGGCCCAC
TTTGCCGCCGGACTGCTGTTCGACAGCAAAGAGCAGGACAAGGACTTCTACCGGAAGTACCCC
ACCATCTATCACCTGAGACTGGCCCTGATGAACGACGACCAGAAGCACGACCTGAGAGAGGTG
TACCTGGCCATCCACCACCTGGTCAAGTACAGAGGCCACTTCCTGATCGAGGGCGACGTGAAA
GCCGACAGCGCCTTTGATGTGCACACCTTCGCCGACGCCATCCAGAGATACGCCGAGAGCAAC
AACTCCGACGAGAACCTGCTGGGCAAGATCGACGAGAAGAAGCTGAGCGCTGCCCTGACCGAT

AAGCACGGCAGCAAAAGCCAGAGAGCCGAGACAGCCGAAACCGCCTTCGACATCCTGGACCTG
CAGTCCAAGAAGCAGATCCAGGCCATCCTGAAGTCCGTCGTGGGCAACCAGGCCAATCTGATG
GCCATTTTTGGCCTGGACAGCAGCGCCATCAGCAAGGACGAGCAGAAGAACTACAAGTTCAGC
TTCGACGACGCCGACATCGATGAGAAGATCGCCGATTCTGAGGCCCTGCTGAGCGATACCGAG
TTCGAGTTCCTGTGCGATCTGAAGGCCGCCTTTGACGGCCTGACACTGAAAATGCTGCTGGGC
GACGACAAGACCGTGTCCGCTGCTATGGTTCGACGGTTCAACGAGCACCAGAAGGACTGGGAG
TACATCAAGAGCCACATCCGGAACGCCAAGAACGCCGGCAATGGCCTGTACGAGAAGTCTAAG
AAGTTCGACGGCATCAACGCCGCCTATCTGGCTCTGCAGTCCGACAACGAGGACGACAGAAAG
AAGGCCAAGAAGATTTTCCAGGACGAGATCAGCTCCGCCGACATTCCCGATGATGTGAAGGCC
GATTTCCTGAAGAAGATTGACGACGATCAGTTCCTGCCTATCCAGCGGACCAAGAACAACGGC
ACAATCCCTCACCAGCTGCACCGGAACGAGCTGGAACAGATCATCGAGAAGCAGGGGATCTAC
TACCCATTCCTGAAGGACACCTACCAAGAGAACAGCCACGAGCTGAACAAAATCACAGCCCTG
ATCAACTTCAGGGTGCCCTACTACGTGGGCCCTCTGGTGGAAGAGGAACAGAAAATCGCCGAC
GACGGCAAGAACATCCCCGATCCTACCAACCACTGGATGGTCCGAAAGTCCAACGACACCATC
ACACCCTGGAACCTGAGCCAGGTGGTCGACCTGGATAAGAGCGGCAGAAGATTCATCGAGCGG
CTGACCGGCACCGATACCTATCTGATCGGAGAGCCCACACTGCCCAAGAACAGCCTGCTGTAC
CAGAAATTCGACGTGCTGCAAGAACTGAACAACATCCGCGTGTCCGGCAGACGGCTGGACATT
AGAGCCAAGCAGGATGCCTTCGAGCACCTGTTCAAGGTGCAGAAAACCGTGTCTGCTACCAAT
CTGAAGGACTTCCTGGTGCAAGCCGGCTACATCAGCGAGGACACCCAGATTGAAGGACTCGCC
GACGTGAACGGAAAGAACTTCAACAACGCCCTGACCACCTACAACTACCTGGTGTCTGTGCTG
GGCCGCGAGTTCGTGGAAAACCCCAGCAACGAGGAACTGCTGGAAGAGATTACCGAGCTGCAG
ACCGTGTTCGAGGACAAGAAGGTGCTGCGGAGACAGCTGGATCAGCTGGACGGACTGAGCGAC
CACAACAGAGAGAAGCTTTCCCGGAAGCACTACACCGGCTGGGGCAGAATCAGCAAGAAGCTG
CTGACCACCAAGATCGTGCAGAACGCCGACAAGATCGATAACCAGACCTTCGATGTGCCCCGG
ATGAACCAGAGCATCATCGACACCCTGTACAACACCAAGATGAACCTGATGGAAATCATCAAC
AATGCCGAGGATGACTTCGGCGTCAGAGCCTGGATCGACAAGCAGAACACCACCGATGGCGAC
GAGCAGGACGTGTACAGCCTGATCGATGAACTGGCTGGCCCCAAAGAGATCAAGCGGGGCATC
GTGCAGTCCTTTAGAATCCTGGACGACATCACCAAGGCCGTGGGCTACGCCCCTAAACGGGTG
TACCTCGAATTTGCCAGAAAGACCCAAGAGAGCCACCTGACCAACAGCCGGAAGAACCAGCTG
AGCACCCTGCTGAAGAATGCCGGCCTGTCTGAGCTGGTCACACAGGTGTCCCAGTATGATGCC
GCCGCTCTGCAGAACGACCGGCTGTATCTTTACTTCCTGCAGCAAGGCAAGGACATGTACTCC
GGCGAGAAGCTGAATCTGGACAACCTGAGCAACTACGACATCGACCACATCATCCCTCAGGCT
TACACCAAGGACAACAGCCTGGACAACAGAGTGCTGGTGTCCAATATCACCAACCGGCGGAAG
TCCGACAGCAGCAACTATCTGCCCGCTCTGATCGATAAGATGCGGCCCTTTTGGAGCGTGCTG
AGCAAGCAGGGGCTGCTGTCTAAGCACAAGTTCGCCAACCTGACCAGAACCAGAGACTTCGAC
GATATGGAAAAAGAGCGGTTTATCGCCCGCAGCCTGGTGGAAACCCGGCAGATCATTAAGAAC
GTGGCCAGCCTGATTGACAGCCACTTCGGCGGAGAGACAAAAGCCGTGGCCATTAGAAGCAGC
CTGACAGCCGACATGCGGAGATACGTGGACATCCCCAAGAACCGGGACATCAACGACTACCAC
CACGCCTTCGATGCCCTGCTGTTTAGCACAGTGGGCCAGTACACCGAGAACAGCGGCCTGATG
AAGAAGGGCCAGCTGTCCGATTCTGCCGGCAACCAGTACAATCGGTACATCAAAGAGTGGATT
CACGCCGCCAGGCTGAACGCACAGTCCCAGAGAGTGAACCCCTTCGGCTTTGTCGTGGGCTCC
ATGAGAAATGCTGCCCCTGGCAAGCTGAACCCCGAGACAGGGGAGATCACCCCAGAGGAAAAC
GCCGACTGGTCTATCGCCGACCTGGACTACCTGCACAAAGTGATGAATTTCCGGAAGATCACC
GTGACCAGGCGGCTGAAGGATCAGAAAGGACAGCTGTACGACGAGAGCAGATACCCCTCCGTG
CTGCACGACGCCAAGTCTAAGGCCAGCATCAACTTTGACAAGCACAAGCCCGTGGACCTGTAC
GGCGGCTTTAGCTCTGCCAAGCCTGCCTATGCCGCACTGATCAAGTTCAAGAACAAGTTCCGG

CTGGTCAACGTGCTGCGGCAGTGGACCTACAGCGACAAGAACTCCGAGGACTATATCCTTGAG
CAGATCAGAGGCAAGTACCCTAAGGCCGAGATGGTGCTGTCTCACATCCCTTACGGCCAGCTG
GTCAAGAAAGATGGCGCCCTGGTCACCATCTCTAGCGCCACAGAGCTGCACAACTTTGAGCAG
CTGTGGCTGCCTCTGGCCGACTACAAGCTGATCAACACACTGCTTAAGACCAAAGAGGACAAC
CTCGTCGATATCCTGCACAACCGGCTGGATCTCCCCGAGATGACAATCGAGAGCGCCTTCTAC
AAAGCCTTCGACTCCATCCTGAGCTTCGCCTTCAACAGATACGCCCTGCACCAGAACGCCCTC
GTGAAACTGCAGGCCCACAGGGACGATTTCAATGCCCTGAACTACGAGGATAAGCAGCAGACC
CTGGAAAGGATTCTGGACGCTCTGCATGCCTCTCCAGCCAGCAGCGACCTGAAGAAAATCAAC
CTGTCCAGCGGCTTCGGCCGGCTGTTTTCCCCTAGCCACTTTACCCTGGCCGACACCGACGAG
TTCATCTTCCAGAGCGTGACCGGCCTGTTCAGCACCCAGAAAACAGTGGCTCAGCTGTATCAA
GAGACAAAGCCCAAGAAGAAGAGGAAGGTG

OkiCas9/NLS protein sequence (SEQ ID NO:16)

MARDYSVGLDIGTSSVGWAAIDNKYHLIRAKSKNLIGVRLFDSAVTAEKRRGYRTTRRRLSRR
HWRLRLLNDIFAGPLTDFGDENFLARLKYSWVHPQDQSNQAHFAAGLLFDSKEQDKDFYRKYP
TIYHLRLALMNDDQKHDLREVYLAIHHLVKYRGHFLIEGDVKADSAFDVHTFADAIQRYAESN
NSDENLLGKIDEKKLSAALTDKHGSKSQRAETAETAFDILDLQSKKQIQAILKSVVGNQANLM
AIFGLDSSAISKDEQKNYKFSFDDADIDEKIADSEALLSDTEFEFLCDLKAAFDGLTLKMLLG
DDKTVSAAMVRRFNEHQKDWEYIKSHIRNAKNAGNGLYEKSKKFDGINAAYLALQSDNEDDRK
KAKKIFQDEISSADIPDDVKADFLKKIDDDQFLPIQRTKNNGTIPHQLHRNELEQIIEKQGIY
YPFLKDTYQENSHELNKITALINFRVPYYVGPLVEEEQKIADDGKNIPDPTNHWMVRKSNDTI
TPWNLSQVVDLDKSGRRFIERLTGTDTYLIGEPTLPKNSLLYQKFDVLQELNNIRVSGRRLDI
RAKQDAFEHLFKVQKTVSATNLKDFLVQAGYISEDTQIEGLADVNGKNFNNALTTYNYLVSVL
GREFVENPSNEELLEEITELQTVFEDKKVLRRQLDQLDGLSDHNREKLSRKHYTGWGRISKKL
LTTKIVQNADKIDNQTFDVPRMNQSIIDTLYNTKMNLMEIINNAEDDFGVRAWIDKQNTTDGD
EQDVYSLIDELAGPKEIKRGIVQSFRILDDITKAVGYAPKRVYLEFARKTQESHLTNSRKNQL
STLLKNAGLSELVTQVSQYDAAALQNDRLYLYFLQQGKDMYSGEKLNLDNLSNYDIDHIIPQA
YTKDNSLDNRVLVSNITNRRKSDSSNYLPALIDKMRPFWSVLSKQGLLSKHKFANLTRTRDFD
DMEKERFIARSLVETRQIIKNVASLIDSHFGGETKAVAIRSSLTADMRRYVDIPKNRDINDYH
HAFDALLFSTVGQYTENSGLMKKGQLSDSAGNQYNRYIKEWIHAARLNAQSQRVNPFGFVVGS
MRNAAPGKLNPETGEITPEENADWSIADLDYLHKVMNFRKITVTRRLKDQKGQLYDESRYPSV
LHDAKSKASINFDKHKPVDLYGGFSSAKPAYAALIKFKNKFRLVNVLRQWTYSDKNSEDYILE
QIRGKYPKAEMVLSHIPYGQLVKKDGALVTISSATELHNFEQLWLPLADYKLINTLLKTKEDN
LVDILHNRLDLPEMTIESAFYKAFDSILSFAFNRYALHQNALVKLQAHRDDFNALNYEDKQQT
LERILDALHASPASSDLKKINLSSGFGRLFSPSHFTLADTDEFIFQSVTGLFSTQKTVAQLYQ
ETKPKKKRKV

BboCas9/NLS DNA sequence (SEQ ID NO:17)

ATGAGCCAGCACCGGCGGTATAGAATCGGCATCGACGTGGGCCTGAATAGCGTTGGACTGGCC
GCCGTGGAAATCGACGCCAACCACGACAATCCTCTGGACGAGATCCCCATCAGCATCCTGAAT
GCCCAGAGCGTGATCCACGATGGCGGAGTGGACCCTGATGAGGCCAAGTCTGCTACAAGCAGA
CGGGCTTCTGCTGGCGTGGCCAGAAGAACAAGACGGCTGCACAAGAGCAAGCGGCAGAGACTG
GCCAAGCTGGACGAGGTGCTGAATGAGCTGGGCTACCCCGTGGAAGATGAGAGCCAGTTTCCA
GCCGGCAGCAACCCCTATATCGCTTGGCAAGTGCGGGCCAAACTGGCCGAGACATTCATCCCC
GACGTGGAAACCCGGAAGCGGATGATCTCTATCGCCATCCGGCACATTGCCCGGCATAGAGGA
TGGCGGAATCCCTACTCTTCTGTGGCCGACGCCGAGCGGATGAGCCATACACCTTCTCCATTC

ATGGTGGAATACGCCAAGAAGCTGGACTTCGAGATCAACGACAGACGGACCAACGGCTTCTAT
CACAGCCCTTGGCAGAGCGTGGACGAGGAAGGCAAGAGACTGAGCAAGAGCGAGCTGGAAAAG
CAGCCCAAGATCGAGGACTGGAACGACAACCCCATCAACGGCAAGACAATCGCCCAGCTGGTC
GTGTCCTCTCTGGAACCCCAGACCAAGATCAGACGGGATCTGACACACGGCCTGCAGACCGAG
AGCACCCTGAATATCCAGACAGAGAAGCTGCACCAGAGCGACTACATCCACGAACTGGAAACC
ATCTTCGAGCGGCAGCACGTGGACCAGACAACCCAAGAACAGCTGCTGGAAGCCACCTTCCAC
ACCAAGAATCCTAAGGCCGTGGGAGCCGCCGCTAAGCTCGTTGGAAAAGATGCCCTGGACAGC
CGGTACTACAGAGCCAGCAGAGCCACACCAGCCTTCGAAGAGTACAGAGTGATGGCCGCCATC
GACACCCTGCGGATTAGAGAGCACGGCACCGAGAGACAGCTGACCACCGACGAGAGAAGAAAG
CTGTTCGACTTCATCAAGGGGCTGCCCAGCAAAAAGACCAAGAACGAGCCCAGCATCAGCTCC
CTGACCTGGGGAGATGTGGCCGATTTTCTGGGCATCCAGCGGATCGATCTGAGAGGCCTGGGC
TCTCTGAAAGACGGCGAACCTGTGTCTGCCAAGCAGCCTCCTGTGATCGAGACAAACGACATC
ATGCAGAAGGCCCCTGATCCAATCGCTGCCTGGTGGTCACAGGCCAACACCAAAGAACGGGAC
AGATTCGTCGAGTTCATGAGCAACGCTGGCGCCATCAAGGACACCTCCGACGAAGTGCGGAAC
ATTGACGCCGAGATCAGCCAGCTGCTCGAAGAACTGACCGGCTCTGAGCTGGAATCCCTGGAT
AAGATCACCCTGACCTCTGGCAGAGCCGCCTACAGCTCTCAGACCCTGAGAAACATCACCAAC
TATATGTACGAGACAGGCTGCGACCTGACCACAGCCAGACAAGAGCTGTACCACGTGGGCAAG
AATTGGGCCCCTCCTGCTCCTCCTATCTACGAGCACACAGGCAACCCCAGCGTGGACAGAACC
TTCAGCATCATCCACAGATGGCTGTGCAACATGCGGGACCAGTACGGCGAGCCCGAGACAGTG
AATATCGAGTACGTCCGCGACGGCTTCAGCAGCACATCTACACAGCTGGCCGAGCAGCGCGAG
CGGGATAGAAGATACGCCGACAACCTGAAGATGCTGAGCAACTACGAGGGCGCCAGCAGCAGA
TCAGATGTGCGGAGAATCAAGGCCCTGCAGAGACAGAACTGCCAGTGCATCTACTGCGGCCGG
ACCATCACCTTCGAGACATGCCAGATGGACCATGTGCTGCCCCGGAAAGGCCCTGGATCCGAT
AGCAAGTTCGAGAACCTGGTGGCCACATGCGGCGAGTGCAACAAGTCCAAGAGCGATACCCTG
TACATGAACTGGGCCAAGACATACCCCAATACCAACCTGCAGGACGTGCTGAGAAGAATCCAA
GAGTGGTCCAAGGACGGCTGGATGACCGACAAAAGATGGCGGCAGTACAAAGAGGCCCTGATC
CTGAGACTGGAAGCTACCGAGAAGCAAGAGCCCCTGGACAATCGGAGCATGGAAAGCGTGTCC
TACATGGCCAGAGAGCTGCGGAACCGGATCTACGGCTTTTACGGCTGGCACGACCAGGACGAC
GCCCTGAAACAAGGCAGACAGAGGGTGTTCGTGTCCAGCGGCAGTATGACAGCCGCTGCCAGA
AGGACCCCTTTCGAGTCCCCACTGATTAAGGGCGCCGATGAGGAAACCTACGAGAGCAGCCTG
CCTTGGCTGGATGGCATGAAGGGCAAGACCAGACTGGATCGGAGACACCATGCCGTGGACGCC
AGCATCATTGCCATGATGAGGCCCCAGATCGTGAAGATCCTGACAGAGGCCCAAGAGATCAGA
AGCGAGCAGCACGACAAGTACCGGAAGGGCCAGACACCTGACTACGTGTGCAAGCGGCGGGAC
TACTGGCGGAATTGGAGAGGCACCCCTGACACCAGAGATGAGGAAGTGTTCAACTACTGGGCT
GGGGAGCAGCTGAGAACCCTGACCGATCTGGTGTCCCAGAAGATGGCCGACGACGAAATCCCC
GTGATCTACCCCACCAGACTGAGACTCGGCAATGGCAGCGCCCACAAGGATACCGTGGTGTCC
ATGATGACCCGGAAAGTGGGCGACGAGCTGAGCATCACCGCCATCAACAAAGCCGAAAGCGGA
GCCCTGTACACAGCCCTGACCAGAGACAGCGACTTCGACTGGAAAACCGGCCTGAGCGCCAAT
CCTAACCGGCGGATCAGAGTGCACGATAAGTGGTTCGAGGCCGACGATACCATCAAGTTTCTG
GAACCTGCCGTGGAAGTGGTGCTGAAGAACAACACCAGAGCCAGAATCGACCCCGAGGCTCTG
GATAAGGTGCACAGCACACTGTACGTGCCCGTCAGAGGCGGAATCGCCGAAGCCGGAAATAGC
ATTCACCACGTGCGGTTCTACAAGATCCCCAAGCTGAACAGCAAGGGCAAGCAGACCGGCAGC
ATCTACGCCATGCTGAGAGTGCTGACCATCGACCTGGCCATGAACCAGTACGACAAAGAGACA
GGCAAGAAGCAGGACCTGTTCACCCTGCCACTGCCTGAAAGCAGCCTGAGCAGAAGATTCAGC
GAGCCCAAACTGCGGCAGGCTCTGATCGATGGCACAGCCGAATATCTCGGATGGGCCGTCGTG
GACGATGAGCTTGAGATCCCCGCCTTCGCCAACGCCAGAATCACAGAGGAACAGGCCATTAAC

GGCAGCTTCACCGACAGACTGCTGCACAGCTTTCCCGGCACACACAAGTTCAGATTCGCCGGC
TTCTCCCGGAACACCGAGATCGCCATTAGACCTGTGCAGCTGGCCTCTGAGGGCCTGATCGAA
ACCGATGAGAACCGGAAGAGACAGCAGCTGCGGCTGACCCAGCCTAACACCGAGTACAGCAAC
AGCATCAAGAACGTGCTGAAGTCCGGCCTGCACCTGAAAGTGAACACCCTGTTCCAGACAGGC
ATCCTGGTCACCAGGGCCAATAGCCAGGGAAAGCAGAGCATCCGGTTCAGCACAGTGGAAGAG
CCCAAGAAGAAGAGGAAGGTG

BboCas9/NLS protein sequence (SEQ ID NO:18)

MSQHRRYRIGIDVGLNSVGLAAVEIDANHDNPLDEIPISILNAQSVIHDGGVDPDEAKSATSR
RASAGVARRTRRLHKSKRQRLAKLDEVLNELGYPVEDESQFPAGSNPYIAWQVRAKLAETFIP
DVETRKRMISIAIRHIARHRGWRNPYSSVADAERMSHTPSPFMVEYAKKLDFEINDRRTNGFY
HSPWQSVDEEGKRLSKSELEKQPKIEDWNDNPINGKTIAQLVVSSLEPQTKIRRDLTHGLQTE
STLNIQTEKLHQSDYIHELETIFERQHVDQTTQEQLLEATFHTKNPKAVGAAAKLVGKDALDS
RYYRASRATPAFEEYRVMAAIDTLRIREHGTERQLTTDERRKLFDFIKGLPSKKTKNEPSISS
LTWGDVADFLGIQRIDLRGLGSLKDGEPVSAKQPPVIETNDIMQKAPDPIAAWWSQANTKERD
RFVEFMSNAGAIKDTSDEVRNIDAEISQLLEELTGSELESLDKITLTSGRAAYSSQTLRNITN
YMYETGCDLTTARQELYHVGKNWAPPAPPIYEHTGNPSVDRTFSIIHRWLCNMRDQYGEPETV
NIEYVRDGFSSTSTQLAEQRERDRRYADNLKMLSNYEGASSRSDVRRIKALQRQNCQCIYCGR
TITFETCQMDHVLPRKGPGSDSKFENLVATCGECNKSKSDTLYMNWAKTYPNTNLQDVLRRIQ
EWSKDGWMTDKRWRQYKEALILRLEATEKQEPLDNRSMESVSYMARELRNRIYGFYGWHDQDD
ALKQGRQRVFVSSGSMTAAARRTPFESPLIKGADEETYESSLPWLDGMKGKTRLDRRHHAVDA
SIIAMMRPQIVKILTEAQEIRSEQHDKYRKGQTPDYVCKRRDYWRNWRGTPDTRDEEVFNYWA
GEQLRTLTDLVSQKMADDEIPVIYPTRLRLGNGSAHKDTVVSMMTRKVGDELSITAINKAESG
ALYTALTRDSDFDWKTGLSANPNRRIRVHDKWFEADDTIKFLEPAVEVVLKNNTRARIDPEAL
DKVHSTLYVPVRGGIAEAGNSIHHVRFYKIPKLNSKGKQTGSIYAMLRVLTIDLAMNQYDKET
GKKQDLFTLPLPESSLSRRFSEPKLRQALIDGTAEYLGWAVVDDELEIPAFANARITEEQAIN
GSFTDRLLHSFPGTHKFRFAGFSRNTEIAIRPVQLASEGLIETDENRKRQQLRLTQPNTEYSN
SIKNVLKSGLHLKVNTLFQTGILVTRANSQGKQSIRFSTVEEPKKKRKV

AceCas9/NLS DNA sequence (SEQ ID NO:19)

ATGGGCGGATCTGAAGTGGGAACCGTGCCTGTGACTTGGAGACTGGGAGTCGATGTGGGCGAG
AGATCCATTGGACTGGCCGCCGTGTCCTACGAAGAGGACAAGCCCAAAGAAATCCTGGCTGCT
GTGTCCTGGATTCACGATGGCGGAGTGGGCGACGAAAGAAGCGGAGCTAGTAGACTGGCCCTG
AGAGGCATGGCCAGAAGGGCTAGACGGCTGCGGAGATTCCGTAGAGCCAGACTGCGCGACCTG
GACATGCTGCTGTCTGAACTCGGATGGACCCCTCTGCCTGACAAGAACGTGTCACCTGTGGAT
GCCTGGCTGGCCAGAAAGAGACTGGCCGAGGAATACGTGGTGGACGAGACAGAGAGAAGAAGG
CTGCTGGGCTACGCCGTGTCTCACATGGCTAGACATAGAGGCTGGCGGAACCCCTGGACCACC
ATCAAGGACCTGAAGAACCTGCCTCAGCCTAGCGACAGCTGGGAGAGAACCAGAGAAAGCCTG
GAAGCCCGGTACTCCGTGTCTCTGGAACCTGGCACAGTTGGACAGTGGGCCGGATACCTGCTG
CAGAGAGCCCCTGGCATCAGACTGAACCCTACACAGCAGAGCGCCGGAAGAAGGGCCGAACTG
TCTAATGCCACCGCCTTCGAGACAAGACTGCGGCAAGAGGATGTGCTGTGGGAGCTGAGATGT
ATCGCCGACGTTCAGGGCCTGCCTGAGGACGTGGTGTCCAATGTGATCGACGCCGTGTTCTGC
CAGAAAAGACCTAGCGTGCCCGCCGAGAGAATCGGCAGAGATCCTCTCGATCCCAGCCAGCTG
AGAGCCAGCAGAGCCTGCCTGGAATTTCAAGAGTACCGGATCGTGGCCGCTGTGGCCAACCTG
AGAATCAGAGATGGCAGCGGCAGCAGACCCCTGAGTCTGGAAGAAAGAAACGCCGTGATCGAG
GCCCTGCTGGCCCAGACAGAAAGAAGCCTCACTTGGAGCGACATTGCCCTGGAAATCCTGAAG

CTGCCCAACGAGAGCGACCTGACCTCTGTGCCTGAAGAGGATGGCCCAAGCAGCCTGGCCTAC
TCTCAGTTCGCCCCTTTCGATGAGACAAGCGCCCGGATCGCCGAGTTTATCGCCAAGAACAGA
CGGAAGATCCCCACATTCGCCCAGTGGTGGCAAGAGCAGGATCGGACCAGTAGAAGCGATCTG
GTGGCTGCCCTGGCCGACAATTCTATTGCCGGCGAGGAAGAACAAGAGCTGCTGGTGCATCTG
CCCGACGCCGAACTTGAAGCTCTGGAAGGACTGGCTCTGCCCTCTGGCAGAGTGGCCTATAGC
AGACTGACACTGAGCGGCCTGACCAGAGTGATGAGAGATGATGGCGTGGACGTGCACAACGCC
CGCAAGACATGCTTCGGAGTGGACGACAATTGGCGGCCTCCACTGCCTGCTCTGCATGAAGCT
ACAGGACACCCCGTGGTGGATAGAAACCTGGCTATCCTGCGGAAGTTCCTGAGCAGCGCCACC
ATGAGATGGGGCCCTCCACAGTCTATCGTGGTGGAACTTGCCAGAGGCGCCAGCGAGAGCAGA
GAAAGGCAGGCCGAAGAAGAAGCCGCTCGGAGAGCCCACAGAAAGGCCAACGACAGAATTAGA
GCCGAACTCAGAGCCTCCGGCCTGAGCGATCCTTCTCCTGCCGATCTTGTTAGAGCCCGGCTG
CTGGAACTGTACGACTGCCACTGTATGTACTGTGGCGCCCCTATCTCCTGGGAGAACAGCGAG
CTGGATCACATCGTGCCCAGAACAGATGGCGGATCCAACAGACACGAGAACCTGGCCATTACA
TGCGGCGCCTGCAACAAAGAAAAAGGCAGAAGGCCCTTCGCCAGCTGGGCCGAGACAAGCAAT
AGAGTGCAGCTGCGGGACGTGATCGACCGGGTGCAGAAGCTGAAGTACAGCGGCAACATGTAC
TGGACCCGGGACGAGTTCAGCCGGTACAAGAAAAGCGTGGTGGCCCGGCTGAAGCGGAGAACC
TCTGATCCTGAAGTGATCCAGAGCATCGAGAGCACCGGCTATGCTGCCGTGGCTCTGAGAGAT
AGACTGCTGAGCTACGGCGAGAAGAATGGCGTGGCACAGGTGGCCGTTTTTAGAGGCGGAGTG
ACAGCCGAGGCCAGAAGATGGCTGGACATCTCCATCGAGCGGCTGTTCAGTAGAGTGGCCATC
TTCGCCCAGAGCACCTCCACCAAGAGGCTGGATAGAAGGCACCACGCCGTGGATGCTGTGGTG
CTGACAACACTGACACCCGGCGTGGCCAAGACACTGGCTGATGCTAGAAGCAGAAGAGTGTCC
GCCGAGTTCTGGCGCAGACCAAGCGACGTGAACAGACACAGCACCGAGGAACCTCAGAGCCCC
GCCTACAGACAGTGGAAAGAGAGCTGTTCTGGCCTGGGCGACCTGCTGATTTCTACCGCCGCC
AGAGATTCTATCGCCGTGGCTGCTCCTCTGAGACTGAGGCCAACAGGCGCACTGCACGAGGAA
ACCCTGAGAGCCTTTAGCGAGCACACAGTGGGAGCCGCTTGGAAGGGCGCTGAGCTGAGAAGA
ATCGTGGAACCCGAAGTGTACGCCGCCTTCCTGGCACTTACAGATCCTGGCGGCAGATTCCTG
AAGGTGTCCCCTAGCGAAGATGTGCTGCCTGCCGACGAGAACAGGCACATTGTGCTGAGCGAC
AGAGTGCTGGGCCCCAGAGACAGAGTGAAACTGTTCCCCGACGACCGGGGCAGCATCAGAGTC
AGAGGTGGCGCAGCCTATATCGCCAGCTTTCACCACGCCAGAGTGTTCAGATGGGGAAGCAGC
CACTCTCCTAGCTTCGCCCTGCTGAGAGTCTCTCTGGCTGATCTGGCTGTGGCTGGCCTGCTT
AGAGATGGGGTCGACGTGTTCACAGCCGAGCTGCCACCTTGGACTCCCGCTTGGAGATATGCC
TCTATCGCCCTGGTCAAGGCCGTGGAAAGCGGCGACGCTAAGCAAGTTGGATGGCTGGTGCCT
GGCGACGAACTGGATTTTGGACCTGAGGGCGTGACAACCGCTGCCGGCGATCTGAGCATGTTC
CTGAAGTACTTTCCCGAGCGGCACTGGGTCGTGACCGGCTTCGAAGATGACAAGAGGATCAAC
CTGAAGCCTGCCTTCCTGTCTGCCGAACAGGCTGAGGTGCTGAGGACTGAGAGAAGCGACAGA
CCCGACACACTGACAGAGGCCGGCGAAATTCTGGCCCAGTTCTTCCCTAGATGTTGGCGGGCC
ACAGTGGCTAAGGTGCTGTGCCATCCTGGCCTGACCGTGATCAGAAGAACAGCCCTGGGACAG
CCTAGGTGGCGGAGAGGACATCTGCCTTATTCATGGCGGCCTTGGAGCGCCGATCCTTGGAGT
GGCGGAACACCTCCCAAGAAGAAGAGGAAGGTG

AceCas9/NLS protein sequence (SEQ ID NO:20)

MGGSEVGTVPVTWRLGVDVGERSIGLAAVSYEEDKPKEILAAVSWIHDGGVGDERSGASRLAL
RGMARRARRLRRFRRARLRDLDMLLSELGWTPLPDKNVSPVDAWLARKRLAEEYVVDETERRR
LLGYAVSHMARHRGWRNPWTTIKDLKNLPQPSDSWERTRESLEARYSVSLEPGTVGQWAGYLL
QRAPGIRLNPTQQSAGRRAELSNATAFETRLRQEDVLWELRCIADVQGLPEDVVSNVIDAVFC
QKRPSVPAERIGRDPLDPSQLRASRACLEFQEYRIVAAVANLRIRDGSGSRPLSLEERNAVIE

ALLAQTERSLTWSDIALEILKLPNESDLTSVPEEDGPSSLAYSQFAPFDETSARIAEFIAKNR
RKIPTFAQWWQEQDRTSRSDLVAALADNSIAGEEEQELLVHLPDAELEALEGLALPSGRVAYS
RLTLSGLTRVMRDDGVDVHNARKTCFGVDDNWRPPLPALHEATGHPVVDRNLAILRKFLSSAT
MRWGPPQSIVVELARGASESRERQAEEEAARRAHRKANDRIRAELRASGLSDPSPADLVRARL
LELYDCHCMYCGAPISWENSELDHIVPRTDGGSNRHENLAITCGACNKEKGRRPFASWAETSN
RVQLRDVIDRVQKLKYSGNMYWTRDEFSRYKKSVVARLKRRTSDPEVIQSIESTGYAAVALRD
RLLSYGEKNGVAQVAVFRGGVTAEARRWLDISIERLFSRVAIFAQSTSTKRLDRRHHAVDAVV
LTTLTPGVAKTLADARSRRVSAEFWRRPSDVNRHSTEEPQSPAYRQWKESCSGLGDLLISTAA
RDSIAVAAPLRLRPTGALHEETLRAFSEHTVGAAWKGAELRRIVEPEVYAAFLALTDPGGRFL
KVSPSEDVLPADENRHIVLSDRVLGPRDRVKLFPDDRGSIRVRGGAAYIASFHHARVFRWGSS
HSPSFALLRVSLADLAVAGLLRDGVDVFTAELPPWTPAWRYASIALVKAVESGDAKQVGWLVP
GDELDFGPEGVTTAAGDLSMFLKYFPERHWVVTGFEDDKRINLKPAFLSAEQAEVLRTERSDR
PDTLTEAGEILAQFFPRCWRATVAKVLCHPGLTVIRRTALGQPRWRRGHLPYSWRPWSADPWS
GGTPPKKKRKV

AheCas9/NLS DNA sequence (SEQ ID NO:21)

ATGGCCTATAGACTGGGCCTCGACATCGGCATCACATCTGTTGGATGGGCCGTCGTGGCCCTG
GAAAAGGATGAGTCTGGACTGAAGCCCGTGCGCATCCAGGATCTGGGCGTCAGAATCTTCGAC
AAGGCCGAGGATAGCAAGACCGGCGCTTCTCTGGCTCTGCCCAGAAGAGAAGCCAGAAGCGCC
AGAAGAAGAACCCGGCGGAGAAGGCACAGACTGTGGCGCGTGAAAAGACTGCTGGAACAGCAC
GGCATCCTGAGCATGGAACAGATCGAGGCCCTGTACGCCCAGAGAACAAGCAGCCCTGATGTG
TATGCCCTGAGAGTGGCCGGCCTGGACAGATGTCTGATCGCCGAAGAGATCGCCCGGGTGCTG
ATTCACATTGCCCACAGAAGAGGCTTCCAGAGCAACAGAAAGAGCGAGATCAAGGACAGCGAC
GCCGGCAAGCTGCTGAAGGCCGTGCAAGAGAACGAGAACCTGATGCAGAGCAAGGGCTACAGA
ACCGTGGCCGAGATGCTGGTGTCTGAGGCCACAAAGACAGACGCCGAGGGAAAGCTGGTGCAC
GGCAAGAAGCACGGCTACGTCAGCAACGTGCGGAACAAGGCCGGCGAGTACAGACACACAGTG
TCCAGACAGGCCATCGTGGACGAAGTGCGGAAGATTTTCGCCGCTCAGAGAGCCCTGGGCAAC
GACGTGATGAGCGAGGAACTGGAAGATAGCTACCTGAAGATCCTGTGCAGCCAGCGGAACTTC
GATGATGGCCCTGGCGGCGATTCTCCTTATGGACACGGAAGCGTTAGCCCCGACGGCGTCAGA
CAGAGCATCTACGAGAGAATGGTCGGAAGCTGCACCTTCGAGACAGGCGAGAAGAGAGCCCCT
AGAAGCAGCTACAGCTTCGAGCGGTTTCAGCTGCTGACCAAGGTGGTCAACCTGCGGATCTAC
CGGCAGCAAGAGGATGGCGGCAGATACCCTTGTGAACTGACCCAGACCGAGCGGGCCAGAGTG
ATCGATTGTGCCTACGAGCAGACCAAGATCACCTACGGAAAGCTGAGAAAGCTGCTGGACATG
AAGGACACCGAGAGCTTTGCCGGCCTGACCTACGGCCTGAACAGAAGCAGAAACAAGACCGAG
GACACCGTGTTCGTGGAAATGAAGTTCTACCACGAAGTCCGCAAGGCCCTGCAGAGAGCCGGG
GTTTTCATTCAGGACCTGAGCATCGAGACACTGGACCAGATCGGCTGGATTCTGAGCGTGTGG
AAGTCCGACGACAACCGGCGGAAGAAGCTGTCTACACTGGGCCTGAGCGACAACGTGATCGAA
GAACTGCTGCCCCTGAACGGCTCCAAGTTTGGCCACCTGAGCCTGAAGGCCATCAGAAAGATC
CTGCCTTTCCTGGAAGATGGGTACAGCTACGACGTGGCCTGTGAACTGGCCGGCTATCAGTTT
CAGGGCAAGACAGAGTACGTGAAGCAGCGGCTGCTGCCTCCACTTGGAGAAGGCGAAGTGACA
AACCCCGTTGTGCGCAGAGCACTGAGCCAGGCCATCAAGGTTGTGAACGCCGTGATCAGAAAG
CACGGCAGCCCAGAGAGCATCCACATCGAACTGGCCAGAGAGCTGAGCAAGAACCTGGACGAG
CGGAGAAAGATCGAGAAGGCCCAGAAAGAAAATCAGAAGAACAACGAGCAGATTAAGGACGAG
ATCCGCGAGATCCTGGGATCCGCCCATGTGACCGGAAGAGACATCGTGAAGTACAAGCTGTTC
AAACAGCAACAAGAGTTCTGCATGTACAGCGGCGAGAAGCTGGACGTGACCAGACTGTTTGAG
CCTGGCTATGCCGAGGTGGACCACATCATCCCTTACGGCATCAGCTTCGACGACTCCTACGAC

AACAAGGTGCTGGTTAAGACCGAGCAGAACCGGCAGAAGGGCAATAGAACCCCTCTGGAATAC
CTGCGGGACAAGCCTGAGCAGAAGGCCAAGTTTATCGCCCTGGTGGAATCTATCCCTCTGAGC
CAGAAAAAGAAAAACCACCTCCTGATGGACAAGCGGGCCATCGACCTGGAACAAGAGGGCTTC
AGAGAGCGGAACCTGAGCGATACCCGGTACATCACACGGGCCCTGATGAACCACATCCAGGCT
TGGCTGCTGTTCGACGAGACAGCCAGCACCAGATCCAAGAGGGTCGTGTGTGTGAATGGCGCC
GTGACCGCCTACATGAGAGCTAGATGGGGCCTGACAAAGGATAGAGATGCCGGCGATAAGCAC
CACGCCGCTGATGCTGTGGTGGTGGCCTGTATCGGAGACAGCCTGATCCAGAGAGTGACCAAA
TACGACAAGTTCAAGCGGAACGCCCTGGCCGACCGGAACAGATATGTGCAGCAGGTTTCCAAG
AGCGAGGGCATCACCCAGTACGTGGACAAAGAAACCGGCGAGGTGTTCACCTGGGAGTCCTTC
GATGAGCGGAAGTTCCTGCCTAACGAGCCCCTGGAACCTTGGCCATTCTTCAGGGATGAGCTG
CTGGCCAGACTGAGCGACGACCCCTCCAAGAACATCAGAGCCATCGGCCTGCTGACCTACAGC
GAGACTGAGCAGATCGATCCCATCTTCGTGTCCAGAATGCCCACCAGAAAAGTGACCGGCGCA
GCCCACAAAGAGACAATCAGATCCCCACGGATCGTGAAGGTGGACGATAACAAGGGCACCGAG
ATCCAGGTGGTGGTGTCTAAGGTGGCCCTGACCGAGCTGAAGCTGACCAAAGACGGCGAAATC
AAGGATTACTTCAGGCCCGAGGACGACCCCAGACTGTACAACACCCTGAGAGAACGGCTGGTG
CAGTTCGGCGGAGATGCCAAGGCCGCCTTCAAAGAACCCGTGTACAAGATCAGCAAGGACGGC
TCTGTGCGGACCCCTGTGCGGAAAGTGAAGATTCAAGAGAAGCTGACACTGGGCGTGCCAGTG
CATGGCGGAAGAGGAATTGCCGAGAATGGCGGCATGGTCCGAATCGACGTGTTCGCCAAAGGC
GGCAAGTACTACTTCGTGCCCATCTACGTGGCCGACGTGCTGAAGAGAGAGCTGCCCAACAGA
CTGGCCACCGCTCACAAGCCTTACAGCGAATGGCGCGTGGTGGACGACAGCTACCAGTTCAAG
TTCTCTCTGTACCCCAACGATGCCGTGATGATCAAGCCCAGCAGAGAGGTGGACATCACCTAC
AAGGACCGGAAAGAGCCCGTCGGCTGCCGGATCATGTACTTTGTGTCCGCCAATATCGCCAGC
GCCTCCATCAGCCTGAGAACCCACGATAACTCCGGCGAGCTGGAAGGACTGGGCATCCAAGGA
CTGGAAGTGTTCGAGAAATACGTCGTGGGCCCTCTGGGCGACACACACCCTGTGTACAAAGAA
CGGCGGATGCCCTTCAGAGTGGAACGGAAGATGAACCCCAAGAAGAAGAGGAAGGTG

AheCas9/NLS protein sequence (SEQ ID NO:22)

MAYRLGLDIGITSVGWAVVALEKDESGLKPVRIQDLGVRIFDKAEDSKTGASLALPRREARSA
RRRTRRRRHRLWRVKRLLEQHGILSMEQIEALYAQRTSSPDVYALRVAGLDRCLIAEEIARVL
IHIAHRRGFQSNRKSEIKDSDAGKLLKAVQENENLMQSKGYRTVAEMLVSEATKTDAEGKLVH
GKKHGYVSNVRNKAGEYRHTVSRQAIVDEVRKIFAAQRALGNDVMSEELEDSYLKILCSQRNF
DDGPGGDSPYGHGSVSPDGVRQSIYERMVGSCTFETGEKRAPRSSYSFERFQLLTKVVNLRIY
RQQEDGGRYPCELTQTERARVIDCAYEQTKITYGKLRKLLDMKDTESFAGLTYGLNRSRNKTE
DTVFVEMKFYHEVRKALQRAGVFIQDLSIETLDQIGWILSVWKSDDNRRKKLSTLGLSDNVIE
ELLPLNGSKFGHLSLKAIRKILPFLEDGYSYDVACELAGYQFQGKTEYVKQRLLPPLGEGEVT
NPVVRRALSQAIKVVNAVIRKHGSPESIHIELARELSKNLDERRKIEKAQKENQKNNEQIKDE
IREILGSAHVTGRDIVKYKLFKQQQEFCMYSGEKLDVTRLFEPGYAEVDHIIPYGISFDDSYD
NKVLVKTEQNRQKGNRTPLEYLRDKPEQKAKFIALVESIPLSQKKKNHLLMDKRAIDLEQEGF
RERNLSDTRYITRALMNHIQAWLLFDETASTRSKRVVCVNGAVTAYMRARWGLTKDRDAGDKH
HAADAVVVACIGDSLIQRVTKYDKFKRNALADRNRYVQQVSKSEGITQYVDKETGEVFTWESF
DERKFLPNEPLEPWPFFRDELLARLSDDPSKNIRAIGLLTYSETEQIDPIFVSRMPTRKVTGA
AHKETIRSPRIVKVDDNKGTEIQVVVSKVALTELKLTKDGEIKDYFRPEDDPRLYNTLRERLV
QFGGDAKAAFKEPVYKISKDGSVRTPVRKVKIQEKLTLGVPVHGGRGIAENGGMVRIDVFAKG
GKYYFVPIYVADVLKRELPNRLATAHKPYSEWRVVDDSYQFKFSLYPNDAVMIKPSREVDITY
KDRKEPVGCRIMYFVSANIASASISLRTHDNSGELEGLGIQGLEVFEKYVVGPLGDTHPVYKE
RRMPFRVERKMNPKKKRKV

WsuCas9/NLS DNA sequence (SEQ ID NO:23)

ATGCTGGTGTCCCCTATCTCTGTGGATCTCGGCGGCAAGAATACCGGCTTCTTCAGCTTCACC
GACAGCCTGGACAATAGCCAGAGCGGCACCGTGATCTACGACGAGAGCTTCGTGCTGAGCCAA
GTGGGCAGAAGAAGCAAGCGGCACAGCAAGCGGAACAACCTGAGAAACAAGCTGGTCAAGCGG
CTGTTCCTGCTGATCCTGCAAGAGCACCACGGCCTGAGCATCGACGTTCTGCCCGATGAGATC
CGGGGCCTGTTCAACAAGAGAGGCTACACCTACGCCGGCTTCGAGCTGGACGAGAAGAAGAAG
GACGCCCTGGAAAGCGATACCCTGAAAGAGTTCCTGAGCGAGAAGCTGCAGTCCATCGACAGA
GACAGCGACGTGGAAGATTTCCTGAACCAGATCGCCAGCAACGCCGAGAGCTTTAAGGACTAC
AAGAAAGGCTTCGAGGCCGTGTTCGCCAGCGCCACACACAGCCCCAACAAGAAGCTGGAACTG
AAGGACGAGCTGAAGTCCGAGTACGGCGAGAACGCCAAAGAACTGCTGGCCGGCCTGAGAGTG
ACCAAAGAGATCCTGGACGAGTTCGACAAGCAAGAGAACCAGGGCAACCTGCCTCGGGCCAAG
TACTTTGAGGAACTGGGCGAGTATATCGCCACCAACGAGAAAGTCAAGAGCTTCTTCGACAGC
AACAGCCTGAAGCTGACCGACATGACCAAGCTGATCGGCAACATCAGCAACTACCAGCTGAAA
GAGCTGCGGCGGTACTTCAACGACAAAGAGATGGAAAAGGGCGACATCTGGATTCCCAACAAG
CTGCACAAGATCACCGAGAGATTTGTGCGGAGCTGGCACCCCAAGAACGACGCCGATAGACAG
AGAAGGGCCGAGCTGATGAAGGACCTGAAGTCCAAAGAAATCATGGAACTGCTGACCACCACC
GAGCCTGTGATGACAATCCCTCCTTACGACGACATGAACAACAGAGGCGCCGTGAAGTGTCAG
ACCCTGCGGCTGAATGAGGAATACCTGGACAAACATCTGCCCAACTGGCGGGATATCGCCAAG
AGACTGAACCACGGCAAGTTCAACGACGACCTGGCCGACTCTACCGTGAAGGGCTACAGCGAG
GATAGCACCCTGCTGCACAGACTGCTGGACACCTCTAAAGAGATCGACATCTACGAGCTGCGG
GGCAAGAAGCCCAACGAGCTGCTGGTTAAGACACTGGGCCAGAGCGACGCCAACAGACTGTAT
GGCTTCGCCCAGAACTACTATGAGCTGATCCGGCAGAAAGTGCGCGCTGGCATTTGGGTGCCC
GTGAAGAACAAGGATGACTCCCTGAACCTGGAAGATAACTCCAACATGCTGAAGCGGTGCAAC
CACAATCCTCCACACAAGAAGAATCAGATCCACAACCTGGTGGCCGGCATCCTGGGAGTGAAA
CTGGATGAGGCCAAGTTCGCCGAGTTCGAGAAAGAGCTTTGGAGCGCCAAAGTGGGCAACAAG
AAACTGAGCGCCTACTGCAAGAACATCGAGGAACTGAGAAAGACCCACGGCAACACCTTCAAG
ATCGATATAGAGGAACTGCGCAAGAAGGACCCCGCCGAGCTGTCCAAAGAGGAAAAGGCCAAG
CTGAGACTGACCGACGACGTGATCCTGAATGAGTGGTCCCAGAAGATCGCCAACTTCTTTGAC
ATCGACGACAAGCACCGGCAGCGGTTCAACAACCTGTTCAGCATGGCCCAGCTGCACACAGTG
ATCGACACACCCAGAAGCGGCTTCAGCTCTACCTGCAAAAGATGCACCGCCGAGAACAGGTTC
AGAAGCGAGACAGCCTTCTACAACGACGAGACAGGCGAGTTCCACAAGAAGGCCACAGCCACC
TGTCAGAGACTGCCCGCTGATACCCAGAGGCCTTTCAGCGGAAAGATCGAGCGGTACATCGAC
AAGCTGGGATACGAGCTGGCCAAGATCAAGGCTAAAGAACTGGAAGGCATGGAAGCTAAAGAA
ATCAAGGTGCCCATCATCCTGGAACAGAACGCCTTCGAGTACGAGGAAAGCCTGCGGAAGTCT
AAGACCGGATCCAACGACAGAGTGATCAACTCCAAGAAAGACCGCGACGGAAAGAAACTGGCC
AAGGCCAAAGAGAACGCCGAGGACAGGCTGAAGGACAAGGACAAGCGGATCAAGGCCTTCAGC
AGCGGCATCTGCCCTTACTGCGGAGATACCATCGGAGATGACGGCGAGATCGACCACATCCTG
CCTAGAAGCCACACACTGAAAATCTACGGGACCGTGTTCAACCCCGAGGGCAATCTGATCTAC
GTGCACCAGAAGTGCAACCAGGCCAAAGCCGACAGCATCTACAAGCTGAGCGATATCAAGGCC
GGCGTGTCAGCCCAGTGGATTGAAGAACAGGTGGCCAACATTAAGGGGTACAAGACCTTCAGC
GTGCTGTCCGCCGAACAGCAGAAGGCCTTTAGATACGCCCTGTTCCTCCAGAACGACAACGAG
GCCTACAAAAAGGTGGTGGACTGGCTGCGGACCGACCAGTCTGCTAGAGTGAACGGCACACAG
AAGTACCTGGCCAAAAAGATCCAAGAGAAGCTCACCAAGATGCTGCCTAACAAGCACCTGAGC
TTCGAGTTCATCCTGGCCGATGCCACCGAGGTGTCAGAGCTGAGAAGGCAGTACGCCAGACAG
AACCCTCTGCTGGCTAAGGCCGAGAAGCAGGCCCCTTCTTCTCACGCCATTGATGCCGTGATG

GCCTTCGTGGCCAGATACCAGAAGGTGTTCAAGGACGGCACCCCTCCTAACGCCGATGAGGTG
GCAAAACTGGCTATGCTGGACAGCTGGAACCCCGCCTCTAATGAGCCTCTGACAAAGGGCCTG
TCCACGAACCAGAAAATCGAGAAGATGATCAAGAGCGGCGACTACGGCCAGAAAAACATGAGA
GAGGTGTTCGGCAAGTCCATCTTCGGAGAGAATGCCATCGGCGAGAGATACAAGCCCATCGTG
GTTCAAGAAGGCGGCTACTACATCGGCTACCCCGCCACAGTGAAAAAGGGCTACGAACTGAAG
AACTGCAAGGTGGTCACCAGCAAGAACGATATTGCCAAGCTGGAAAAGATCATCAAGAACCAG
GACCTGATCTCTCTGAAAGAGAATCAGTACATCAAAATCTTCTCCATCAACAAGCAGACCATC
AGCGAGCTGAGCAACCGCTACTTCAACATGAATTACAAGAACCTGGTCGAGCGGGACAAAGAA
ATTGTGGGACTGCTTGAGTTTATCGTCGAGAACTGCCGGTACTACACCAAGAAAGTGGACGTG
AAGTTCGCCCCTAAGTACATCCACGAGACAAAGTACCCCTTCTACGATGACTGGCGGAGATTC
GACGAGGCCTGGCGGTATCTGCAAGAAAACCAGAACAAGACCAGCTCCAAGGACCGCTTCGTG
ATCGATAAGAGCAGCCTGAACGAGTACTACCAGCCAGACAAGAATGAGTACAAGCTGGACGTG
GACACCCAGCCTATCTGGGACGACTTCTGCCGGTGGTACTTCCTGGACAGATACAAGACCGCC
AACGACAAGAAGTCCATCCGCATCAAGGCCCGCAAGACATTCTCCCTGCTGGCTGAGTCTGGC
GTGCAGGGCAAAGTGTTCCGGGCCAAGAGAAAGATCCCTACCGGCTACGCCTATCAGGCCCTG
CCTATGGACAACAACGTGATCGCTGGCGATTACGCCAACATTCTGCTGGAAGCCAACAGCAAG
ACCCTGAGCCTGGTGCCTAAGAGCGGCATCAGCATTGAGAAGCAGCTGGACAAAAAGCTCGAC
GTCATCAAAAAGACCGACGTGCGCGGCCTGGCAATCGACAACAACTCCTTCTTCAACGCCGAC
TTCGACACACACGGCATCCGGCTGATCGTGGAAAACACCAGCGTGAAAGTGGGAAACTTCCCC
ATCAGCGCCATCGATAAGTCCGCCAAGCGGATGATCTTCAGAGCCCTGTTTGAGAAAGAGAAG
GGGAAGCGCAAGAAAAAGACCACCATCAGCTTCAAAGAAAGCGGCCCTGTGCAGGACTACCTC
AAGGTGTTCCTGAAAAAGATCGTGAAGATCCAGCTGAGAACCGACGGCTCCATCTCCAACATC
GTCGTGCGGAAGAATGCCGCCGATTTCACCCTGAGCTTTAGAAGCGAGCACATCCAGAAACTG
CTGAAGCCCAAGAAGAAGAGGAAGGTG

WsuCas9/NLS protein sequence (SEQ ID NO:24)

MLVSPISVDLGGKNTGFFSFTDSLDNSQSGTVIYDESFVLSQVGRRSKRHSKRNNLRNKLVKR
LFLLILQEHHGLSIDVLPDEIRGLFNKRGYTYAGFELDEKKKDALESDTLKEFLSEKLQSIDR
DSDVEDFLNQIASNAESFKDYKKGFEAVFASATHSPNKKLELKDELKSEYGENAKELLAGLRV
TKEILDEFDKQENQGNLPRAKYFEELGEYIATNEKVKSFFDSNSLKLTDMTKLIGNISNYQLK
ELRRYFNDKEMEKGDIWIPNKLHKITERFVRSWHPKNDADRQRRAELMKDLKSKEIMELLTTT
EPVMTIPPYDDMNNRGAVKCQTLRLNEEYLDKHLPNWRDIAKRLNHGKFNDDLADSTVKGYSE
DSTLLHRLLDTSKEIDIYELRGKKPNELLVKTLGQSDANRLYGFAQNYYELIRQKVRAGIWVP
VKNKDDSLNLEDNSNMLKRCNHNPPHKKNQIHNLVAGILGVKLDEAKFAEFEKELWSAKVGNK
KLSAYCKNIEELRKTHGNTFKIDIEELRKKDPAELSKEEKAKLRLTDDVILNEWSQKIANFFD
IDDKHRQRFNNLFSMAQLHTVIDTPRSGFSSTCKRCTAENRFRSETAFYNDETGEFHKKATAT
CQRLPADTQRPFSGKIERYIDKLGYELAKIKAKELEGMEAKEIKVPIILEQNAFEYEESLRKS
KTGSNDRVINSKKDRDGKKLAKAKENAEDRLKDKDKRIKAFSSGICPYCGDTIGDDGEIDHIL
PRSHTLKIYGTVFNPEGNLIYVHQKCNQAKADSIYKLSDIKAGVSAQWIEEQVANIKGYKTFS
VLSAEQQKAFRYALFLQNDNEAYKKVVDWLRTDQSARVNGTQKYLAKKIQEKLTKMLPNKHLS
FEFILADATEVSELRRQYARQNPLLAKAEKQAPSSHAIDAVMAFVARYQKVFKDGTPPNADEV
AKLAMLDSWNPASNEPLTKGLSTNQKIEKMIKSGDYGQKNMREVFGKSIFGENAIGERYKPIV
VQEGGYYIGYPATVKKGYELKNCKVVTSKNDIAKLEKIIKNQDLISLKENQYIKIFSINKQTI
SELSNRYFNMNYKNLVERDKEIVGLLEFIVENCRYYTKKVDVKFAPKYIHETKYPFYDDWRRF
DEAWRYLQENQNKTSSKDRFVIDKSSLNEYYQPDKNEYKLDVDTQPIWDDFCRWYFLDRYKTA
NDKKSIRIKARKTFSLLAESGVQGKVFRAKRKIPTGYAYQALPMDNNVIAGDYANILLEANSK

TLSLVPKSGISIEKQLDKKLDVIKKTDVRGLAIDNNSFFNADFDTHGIRLIVENTSVKVGNFP
ISAIDKSAKRMIFRALFEKEKGKRKKKTTISFKESGPVQDYLKVFLKKIVKIQLRTDGSISNI
VVRKNAADFTLSFRSEHIQKLLKPKKKRKV

NsaCas9/NLS DNA sequence (SEQ ID NO:25)

ATGAAGAAGATCCTGGGCGTCGACCTGGGCATCACCAGCTTTGGATACGCCATCCTGCAAGAG
ACAGGCAAGGACCTGTACAGATGCCTGGACAACAGCGTGGTCATGCGGAACAACCCCTACGAC
GAGAAGTCTGGCGAGAGCAGCCAGAGCATCCGCAGCACCCAGAAATCCATGCGGCGGCTGATC
GAGAAGCGGAAGAAACGGATCAGATGCGTGGCCCAGACAATGGAACGCTACGGCATCCTGGAC
TACTCCGAGACAATGAAGATCAACGACCCCAAGAACAACCCGATCAAGAACAGATGGCAGCTG
AGAGCCGTGGACGCCTGGAAAAGACCTCTGAGCCCTCAAGAGCTGTTCGCCATCTTTGCCCAC
ATGGCCAAGCACCGGGGCTACAAGTCTATCGCCACCGAGGACCTGATCTACGAGCTGGAACTG
GAACTCGGCCTGAACGACCCTGAGAAAGAGTCCGAGAAGAAGGCCGACGAGCGGAGACAGGTG
TACAACGCCCTGAGACACCTGGAAGAACTGCGGAAGAAGTACGGCGGCGAGACAATCGCCCAG
ACCATCCACAGAGCTGTGGAAGCCGGCGACCTGCGGAGCTACAGAAACCACGACGACTACGAG
AAGATGATCCGCAGAGAGGACATCGAGGAAGAGATTGAGAAGGTCCTGCTGCGGCAGGCTGAA
CTGGGAGCACTTGGACTGCCTGAGGAACAGGTGTCCGAGCTGATCGATGAGCTGAAGGCCTGC
ATCACCGACCAAGAGATGCCCACCATCGACGAGAGCCTGTTCGGCAAGTGCACCTTCTACAAG
GACGAGCTGGCCGCTCCTGCCTACAGCTACCTGTACGACCTGTACCGGCTGTACAAGAAGCTG
GCCGACCTGAACATCGACGGCTACGAAGTGACCCAAGAGGACCGCGAGAAAGTGATCGAGTGG
GTCGAGAAAAAGATCGCCCAGGGCAAGAACCTGAAGAAAATCACCCACAAGGACCTCCGGAAG
ATCCTCGGACTGGCCCCTGAGCAGAAGATTTTCGGCGTCGAGGACGAGAGAATCGTCAAGGGA
AAGAAAGAACCCCGGACCTTCGTGCCCTTCTTCTTCCTGGCCGATATCGCCAAGTTCAAAGAA
CTGTTTGCCAGCATCCAGAAGCACCCCGACGCTCTGCAGATTTTCAGAGAACTGGCCGAGATC
CTGCAGCGGAGCAAGACACCTCAAGAGGCCCTGGATAGACTGAGAGCCCTGATGGCCGGCAAG
GGCATCGACACCGATGACAGAGAGCTGCTGGAACTCTTCAAGAACAAGCGGAGCGGCACAAGA
GAGCTGAGCCACCGCTATATCCTGGAAGCCCTGCCTCTGTTCCTGGAAGGCTATGACGAGAAA
GAGGTGCAGAGAATCCTGGGCTTTGACGACCGCGAGGACTACAGCAGATACCCCAAGAGCCTG
CGGCATCTGCACCTGAGAGAGGGCAACCTGTTCGAGAAAGAAGAGAATCCCATCAACAACCAC
GCCGTGAAGTCCCTGGCTTCTTGGGCCCTGGGACTGATCGCTGACCTGTCTTGGAGATACGGC
CCCTTCGATGAGATCATCCTGGAAACCACCAGGGACGCCCTGCCTGAGAAGATCCGGAAAGAA
ATCGACAAGGCCATGCGCGAGAGAGAGAAAGCCCTGGACAAGATCATCGGCAAGTACAAGAAA
GAGTTCCCCAGCATCGACAAGCGGCTGGCCAGAAAGATTCAGCTGTGGGAGAGACAGAAAGGC
CTCGATCTGTACTCCGGCAAAGTGATCAACCTGAGCCAGCTGCTCGATGGATCCGCCGACATC
GAGCACATCGTGCCTCAGTCTCTCGGCGGCCTGAGCACCGACTACAATACCATCGTGACCCTG
AAGTCCGTGAACGCCGCCAAGGGCAATAGACTGCCTGGCGATTGGCTGGCCGGAAATCCCGAC
TACAGAGAACGGATCGGCATGCTGTCTGAGAAGGGCCTGATCGACTGGAAGAAGAGGAAGAAC
CTGCTGGCCCAGAGCCTGGACGAAATCTACACCGAGAACACCCACAGCAAAGGCATCCGGGCC
ACAAGCTACCTGGAAGCTCTGGTTGCCCAGGTGCTGAAGCGGTACTACCCATTTCCTGATCCT
GAGCTGCGCAAGAATGGCATCGGCGTGCGGATGATCCCCGGAAAAGTGACCAGCAAGACCAGA
AGCCTGCTGGGAATCAAGAGCAAGAGCCGCGAGACAAACTTCCACCACGCCGAGGATGCCCTG
ATTCTGAGCACACTGACCAGAGGCTGGCAGAACCGGCTGCACAGAATGCTGAGAGACAACTAC
GGCAAGAGCGAGGCCGAGCTGAAAGAACTCTGGAAAAAGTACATGCCCCACATCGAGGGCCTG
ACACTGGCCGACTATATCGATGAGGCCTTCCGGCGGTTCATGAGCAAGGGCGAAGAGTCCCTG
TTCTACCGGGACATGTTCGACACCATCCGGTCCATCAGCTACTGGGTCGACAAGAAGCCTCTG
AGCGCCAGCAGCCACAAAGAAACCGTGTACAGCAGCAGACACGAGGTGCCCACACTGAGGAAA

AACATTCTGGAAGCCTTCGACAGCCTGAACGTGATCAAGGACCGGCACAAGCTGACCACCGAA
GAGTTCATGAAGCGCTACGACAAAGAGATCCGGCAGAAGCTGTGGCTGCACCGCATCGGCAAC
ACCAACGACGAGTCTTACCGCGCCGTGGAAGAGAGAGCCACACAGATTGCCCAGATCCTGACC
AGATACCAGCTCATGGACGCCCAGAATGACAAAGAAATTGATGAGAAGTTTCAGCAGGCCCTG
AAAGAGCTGATCACAAGCCCCATCGAAGTGACTGGCAAGCTGCTGCGGAAAATGAGATTCGTG
TACGACAAGCTGAACGCCATGCAGATCGACAGAGGCCTGGTGGAAACCGACAAGAACATGCTG
GGCATCCACATCAGCAAGGGCCCCAATGAGAAGCTGATCTTCAGACGGATGGACGTGAACAAC
GCCCACGAGCTGCAAAAAGAACGCAGCGGAATCCTGTGCTACCTGAACGAGATGCTGTTCATC
TTCAACAAGAAGGGGCTGATTCACTACGGCTGCCTGCGGTCTTACCTCGAAAAAGGCCAGGGC
AGCAAGTATATCGCCCTGTTCAACCCTCGGTTCCCCGCCAATCCTAAGGCTCAGCCTAGCAAG
TTCACCAGCGACAGCAAGATCAAGCAAGTCGGCATCGGCAGCGCCACCGGAATCATTAAGGCC
CACCTGGATCTGGATGGCCACGTGCGCTCTTATGAGGTGTTCGGAACACTGCCCGAGGGCAGC
ATCGAGTGGTTCAAAGAGGAAAGCGGCTACGGCAGAGTGGAAGATGACCCTCACCACCCCAAG
AAGAAGAGGAAGGTG

NsaCas9/NLS protein sequence (SEQ ID NO:26)

MKKILGVDLGITSFGYAILQETGKDLYRCLDNSVVMRNNPYDEKSGESSQSIRSTQKSMRRLI
EKRKKRIRCVAQTMERYGILDYSETMKINDPKNNPIKNRWQLRAVDAWKRPLSPQELFAIFAH
MAKHRGYKSIATEDLIYELELELGLNDPEKESEKKADERRQVYNALRHLEELRKKYGGETIAQ
TIHRAVEAGDLRSYRNHDDYEKMIRREDIEEEIEKVLLRQAELGALGLPEEQVSELIDELKAC
ITDQEMPTIDESLFGKCTFYKDELAAPAYSYLYDLYRLYKKLADLNIDGYEVTQEDREKVIEW
VEKKIAQGKNLKKITHKDLRKILGLAPEQKIFGVEDERIVKGKKEPRTFVPFFFLADIAKFKE
LFASIQKHPDALQIFRELAEILQRSKTPQEALDRLRALMAGKGIDTDDRELLELFKNKRSGTR
ELSHRYILEALPLFLEGYDEKEVQRILGFDDREDYSRYPKSLRHLHLREGNLFEKEENPINNH
AVKSLASWALGLIADLSWRYGPFDEIILETTRDALPEKIRKEIDKAMREREKALDKIIGKYKK
EFPSIDKRLARKIQLWERQKGLDLYSGKVINLSQLLDGSADIEHIVPQSLGGLSTDYNTIVTL
KSVNAAKGNRLPGDWLAGNPDYRERIGMLSEKGLIDWKKRKNLLAQSLDEIYTENTHSKGIRA
TSYLEALVAQVLKRYYPFPDPELRKNGIGVRMIPGKVTSKTRSLLGIKSKSRETNFHHAEDAL
ILSTLTRGWQNRLHRMLRDNYGKSEAELKELWKKYMPHIEGLTLADYIDEAFRRFMSKGEESL
FYRDMFDTIRSISYWVDKKPLSASSHKETVYSSRHEVPTLRKNILEAFDSLNVIKDRHKLTTE
EFMKRYDKEIRQKLWLHRIGNTNDESYRAVEERATQIAQILTRYQLMDAQNDKEIDEKFQQAL
KELITSPIEVTGKLLRKMRFVYDKLNAMQIDRGLVETDKNMLGIHISKGPNEKLIFRRMDVNN
AHELQKERSGILCYLNEMLFIFNKKGLIHYGCLRSYLEKGQGSKYIALFNPRFPANPKAQPSK
FTSDSKIKQVGIGSATGIIKAHLDLDGHVRSYEVFGTLPEGSIEWFKEESGYGRVEDDPHHPK
KKRKV

RsyCas9/NLS DNA sequence (SEQ ID NO:27)

ATGGCCGAGAAGCAGCACAGATGGGGACTCGACATCGGCACCAATTCTATCGGCTGGGCCGTG
ATCGCCCTGATCGAAGGCAGACCTGCTGGACTGGTGGCTACCGGCAGCAGAATCTTTAGCGAC
GGCAGAAACCCCAAGGACGGCAGCTCTCTGGCCGTCGAGAGAAGAGGACCTCGGCAGATGCGG
CGGAGAAGAGACAGATATCTCCGGCGGAGGGACAGATTCATGCAGGCCCTGATCAACGTGGGC
CTGATGCCTGGGGATGCCGCCGCTAGAAAAGCCCTGGTCACCGAGAATCCCTACGTGCTGAGA
CAGAGAGGCCTGGACCAAGCTCTGACCCTGCCTGAATTTGGCAGAGCCCTGTTCCACCTGAAC
CAGCGGAGAGGCTTCCAGAGCAACAGAAAGACCGATCGGGCCACCGCCAAAGAAAGCGGCAAA
GTGAAGAACGCCATTGCCGCCTTCAGAGCCGGCATGGGCAATGCCAGAACAGTGGGAGAAGCC
CTGGCCAGACGACTGGAAGATGGCAGACCAGTGCGGGCCAGAATGGTCGGACAGGGCAAAGAT

GAGCACTACGAGCTGTATATCGCCAGAGAGTGGATCGCCCAAGAGTTCGATGCCCTGTGGGCC
AGCCAGCAGAGATTTCATGCTGAGGTGCTGGCCGACGCCGCCAGAGATAGACTGAGAGCCATC
CTGCTGTTCCAGCGGAAGCTGCTGCCTGTGCCTGTGGGCAAGTGCTTCCTGGAACCTAACCAG
CCTAGAGTGGCCGCTGCTCTGCCTAGCGCTCAGAGATTCAGACTGATGCAAGAGCTGAACCAC
CTGAGAGTGATGACCCTGGCCGACAAGAGAGAGAGGCCCCTGAGCTTCCAAGAGAGAAACGAT
CTGCTGGCCCAGCTGGTGGCCAGACCTAAGTGCGGCTTCGACATGCTGCGGAAGATCGTGTTC
GGCGCCAACAAAGAGGCCTACAGATTCACCATCGAGAGCGAGCGGCGGAAAGAACTGAAGGGC
TGTGATACAGCCGCCAAGCTGGCCAAAGTGAATGCCCTGGGAACTAGATGGCAGGCTCTGTCC
CTGGACGAGCAGGATAGACTCGTGTGCCTGCTGCTGGACGGCGAGAATGATGCTGTGCTGGCT
GATGCCCTGCGGGAACACTATGGACTGACAGACGCCCAGATCGACACACTGCTGGGCCTGTCT
TTTGAGGACGGCCACATGAGACTGGGGAGAAGCGCTCTGCTGAGAGTCCTGGATGCCCTGGAA
TCCGGAAGAGATGAGCAGGGACTGCCCCTGTCCTACGATAAGGCTGTTGTGGCTGCCGGCTAT
CCAGCTCACACAGCCGATCTGGAAAACGGCGAGAGAGATGCACTGCCCTACTACGGCGAGCTG
CTGTGGCGGTATACACAGGATGCCCCTACCGCCAAGAACGACGCCGAGAGAAAGTTCGGCAAG
ATCGCCAATCCTACCGTGCACATCGGCCTGAATCAGCTGAGAAAGCTTGTCAATGCCCTGATC
CAGAGATACGGCAAGCCCGCTCAGATCGTGGTGGAACTGGCCAGAAATCTGAAGGCTGGCCTG
GAAGAGAAAGAGCGGATCAAGAAACAGCAGACCGCCAACCTGGAACGGAACGAGAGAATCCGG
CAGAAGCTGCAGGACGCTGGCGTGCCCGACAACAGAGAAAACCGGCTGCGGATGCGGCTGTTC
GAGGAACTCGGACAAGGCAATGGACTGGGCACCCCTTGCATCTACTCCGGCAGACAGATCAGC
CTGCAGAGACTGTTCAGCAACGACGTGCAGGTCGACCACATCCTGCCTTTCAGCAAGACCCTG
GATGACAGCTTCGCCAACAAGGTGCTCGCCCAGCACGACGCCAACAGATACAAGGGCAACAGA
GGCCCTTTCGAGGCCTTCGGAGCCAACAGAGATGGCTACGCCTGGGACGACATTAGAGCCAGA
GCAGCCGTGCTGCCCCGGAACAAGAGAAACAGATTTGCCGAGACAGCCATGCAGGACTGGCTG
CACAACGAGACTGACTTTCTGGCTCGGCAGCTGACCGATACCGCCTACCTTAGCAGAGTGGCC
AGGCAGTACCTGACCGCCATCTGCAGCAAGGACGACGTGTACGTTAGCCCCGGCAGACTGACT
GCCATGCTGAGAGCTAAGTGGGGCCTGAACAGAGTGCTGGATGGCGTGATGGAAGAACAGGGC
AGACCCGCCGTGAAGAACCGGGATGATCACAGACACCACGCCATCGACGCCGTGGTTATTGGC
GCCACAGATAGAGCCATGCTGCAACAGGTGGCCACACTGGCCGCTAGAGCTAGAGAACAGGAC
GCCGAAAGGCTGATCGGCGACATGCCTACGCCTTGGCCTAATTTCCTTGAGGACGTGCGGGCT
GCCGTGGCCAGATGTGTGGTTTCTCACAAGCCCGACCACGGACCAGAAGGCGGCCTGCATAAC
GATACAGCCTACGGCATTGTGGCCGGACCATTCGAGGATGGCAGATACAGAGTGCGGCACCGG
GTGTCCCTGTTCGATCTGAAACCTGGCGACCTGAGCAACGTCCGCTGTGATGCTCCTCTGCAA
GCCGAGCTGGAACCCATCTTCGAGCAGGACGATGCCAGGGCCAGAGAAGTGGCTCTTACAGCC
CTGGCTGAGCGGTACAGACAGCGGAAAGTGTGGCTGGAAGAACTGATGAGCGTGCTGCCTATC
AGACCCAGAGGCGAGGACGGAAAGACCCTGCCAGATAGCGCTCCTTACAAGGCCTACAAGGGC
GACTCCAACTACTGCTATGAGCTGTTCATCAATGAGCGCGGCAGATGGGATGGCGAGCTGATC
TCTACCTTCCGGGCCAATCAGGCCGCTTACCGGCGGTTCAGAAATGACCCAGCCAGGTTCAGA
AGATACACCGCTGGCGGTAGACCCCTGCTGATGAGACTGTGTATCAACGACTATATCGCCGTG
GGCACAGCCGCCGAGAGGACCATCTTTAGAGTGGTCAAGATGAGCGAGAACAAGATCACTCTG
GCCGAGCACTTCGAAGGCGGAACCCTGAAACAGAGGGATGCCGACAAGGACGATCCCTTCAAG
TATCTGACAAAGAGCCCTGGCGCTCTGCGCGATCTGGGAGCTAGAAGAATCTTCGTGGACCTG
ATCGGCCGCGTGCTGGACCCAGGCATTAAGGGCGATCCCCAAGAAGAAGAGGAAGGTG

RsyCas9/NLS protein sequence (SEQ ID NO:28)

MAEKQHRWGLDIGTNSIGWAVIALIEGRPAGLVATGSRIFSDGRNPKDGSSLAVERRGPRQMR
RRRDRYLRRRDRFMQALINVGLMPGDAAARKALVTENPYVLRQRGLDQALTLPEFGRALFHLN

QRRGFQSNRKTDRATAKESGKVKNAIAAFRAGMGNARTVGEALARRLEDGRPVRARMVGQGKD
EHYELYIAREWIAQEFDALWASQQRFHAEVLADAARDRLRAILLFQRKLLPVPVGKCFLEPNQ
PRVAAALPSAQRFRLMQELNHLRVMTLADKRERPLSFQERNDLLAQLVARPKCGFDMLRKIVF
GANKEAYRFTIESERRKELKGCDTAAKLAKVNALGTRWQALSLDEQDRLVCLLLDGENDAVLA
DALREHYGLTDAQIDTLLGLSFEDGHMRLGRSALLRVLDALESGRDEQGLPLSYDKAVVAAGY
PAHTADLENGERDALPYYGELLWRYTQDAPTAKNDAERKFGKIANPTVHIGLNQLRKLVNALI
QRYGKPAQIVVELARNLKAGLEEKERIKKQQTANLERNERIRQKLQDAGVPDNRENRLRMRLF
EELGQGNGLGTPCIYSGRQISLQRLFSNDVQVDHILPFSKTLDDSFANKVLAQHDANRYKGNR
GPFEAFGANRDGYAWDDIRARAAVLPRNKRNRFAETAMQDWLHNETDFLARQLTDTAYLSRVA
RQYLTAICSKDDVYVSPGRLTAMLRAKWGLNRVLDGVMEEQGRPAVKNRDDHRHHAIDAVVIG
ATDRAMLQQVATLAARAREQDAERLIGDMPTPWPNFLEDVRAAVARCVVSHKPDHGPEGGLHN
DTAYGIVAGPFEDGRYRVRHRVSLFDLKPGDLSNVRCDAPLQAELEPIFEQDDARAREVALTA
LAERYRQRKVWLEELMSVLPIRPRGEDGKTLPDSAPYKAYKGDSNYCYELFINERGRWDGELI
STFRANQAAYRRFRNDPARFRRYTAGGRPLLMRLCINDYIAVGTAAERTIFRVVKMSENKITL
AEHFEGGTLKQRDADKDDPFKYLTKSPGALRDLGARRIFVDLIGRVLDPGIKGDPKKKRKV

CdiCas9/NLS DNA sequence (SEQ ID NO:29)

ATGAAGTACCACGTGGGCATCGACGTGGGCACCTTTTCTGTTGGACTGGCCGCCATCGAAGTG
GACGATGCCGGAATGCCTATCAAGACCCTGAGCCTGGTGTCCCACATCCACGATTCTGGACTG
GACCCCGACGAGATCAAGAGCGCCGTTACAAGACTGGCCAGCAGCGGAATCGCCAGAAGAACC
AGACGGCTGTACCGGCGGAAGAGAAGAAGGCTGCAGCAGCTGGACAAGTTCATCCAGAGACAA
GGCTGGCCCGTGATCGAGCTGGAAGATTACAGCGACCCTCTGTACCCCTGGAAAGTGCGGGCT
GAACTGGCTGCCAGCTATATCGCCGATGAGAAAGAGCGGGGCGAGAAGCTGTCTGTGGCCCTG
AGACACATTGCCAGACACAGAGGATGGCGGAACCCCTACGCCAAGGTGTCCTCTCTGTATCTG
CCTGACGGCCCTAGCGACGCCTTCAAGGCCATCAGAGAGGAAATCAAGAGAGCCAGCGGCCAG
CCTGTGCCTGAAACAGCTACAGTGGGCCAGATGGTCACCCTGTGTGAACTGGGCACCCTGAAG
TTGAGAGGCGAAGGCGGAGTGCTGTCTGCCAGACTCCAGCAGAGCGATTACGCCAGAGAGATC
CAAGAGATTTGCCGGATGCAAGAGATCGGCCAAGAGCTGTACAGAAAGATCATCGATGTGGTG
TTCGCCGCCGAGTCTCCTAAGGGATCTGCCTCTAGCAGAGTGGGCAAAGACCCTCTGCAGCCC
GGCAAGAATAGAGCCCTGAAAGCCTCCGATGCCTTCCAGAGATACCGGATCGCCGCTCTGATC
GGCAACCTGAGAGTTAGAGTGGACGGCGAGAAGAGGATTCTGAGCGTGGAAGAGAAAAACCTG
GTGTTCGACCACCTGGTCAATCTGACCCCTAAGAAAGAACCCGAGTGGGTCACAATCGCCGAG
ATCCTGGGAATCGACAGAGGCCAGCTGATCGGAACCGCCACCATGACAGATGATGGCGAAAGA
GCCGGCGCTCGGCCTCCTACACATGACACCAATCGGAGCATCGTGAACAGCAGAATCGCCCCT
CTGGTGGACTGGTGGAAAACCGCCTCTGCTCTGGAACAGCACGCTATGGTCAAGGCCCTGTCC
AATGCCGAGGTGGACGACTTCGATTCTCCTGAGGGCGCCAAAGTGCAGGCCTTCTTTGCCGAC
CTGGACGACGATGTGCACGCCAAGCTGGATAGCCTGCATCTGCCTGTTGGCAGAGCCGCCTAC
AGCGAGGATACACTTGTGCGGCTGACCAGACGGATGCTGAGTGATGGCGTGGACCTGTACACC
GCCAGACTGCAAGAGTTTGGCATCGAGCCTAGCTGGACCCCTCCAACACCTAGAATCGGAGAG
CCCGTGGGAAACCCCGCTGTGGACAGAGTGCTGAAAACCGTGTCCAGATGGCTGGAAAGCGCC
ACCAAAACATGGGGCGCTCCCGAGAGAGTGATCATCGAACACGTGCGCGAGGGCTTCGTGACC
GAGAAAAGGGCCAGAGAAATGGATGGCGACATGCGGAGAAGGGCCGCCAGAAATGCCAAGCTG
TTCCAAGAAATGCAAGAAAAGCTGAACGTGCAGGGCAAGCCCTCCAGAGCCGACCTTTGGAGA
TACCAGAGCGTGCAGAGACAGAACTGCCAGTGCGCCTACTGTGGCAGCCCTATCACCTTCAGC
AACAGCGAGATGGACCACATCGTGCCTAGAGCCGGCCAGGGATCCACCAACACCAGAGAAAAT
CTGGTGGCCGTGTGCCACAGATGCAACCAGAGCAAGGGCAACACCCCATTCGCCATCTGGGCC

AAGAACACCTCTATCGAGGGCGTGTCCGTGAAAGAAGCCGTGGAAAGAACCAGGCACTGGGTC
ACCGATACCGGCATGAGAAGCACCGACTTCAAGAAATTCACCAAGGCCGTGGTGGAACGGTTC
CAGAGGGCCACAATGGACGAGGAAATTGACGCCCGCAGCATGGAAAGCGTGGCCTGGATGGCC
AATGAGCTGAGAAGTAGAGTGGCCCAGCACTTCGCCAGCCACGGCACAACAGTCAGAGTGTAC
AGAGGCAGCCTGACCGCCGAAGCTCGTAGAGCCTCTGGAATCAGCGGCAAGCTGAAGTTCTTT
GACGGCGTGGGCAAGAGCAGACTGGACAGAAGGCACCACGCCATTGATGCCGCCGTGATCGCC
TTCACCAGCGACTATGTGGCCGAAACACTGGCCGTGCGGAGCAACCTCAAACAGAGCCAGGCT
CACAGACAAGAGGCTCCTCAGTGGCGCGAGTTCACAGGCAAAGATGCCGAACACAGAGCCGCT
TGGAGAGTGTGGTGCCAGAAGATGGAAAAACTGAGCGCCCTGCTGACCGAGGACCTGAGAGAT
GATAGAGTGGTGGTCATGAGCAACGTGCGCCTGAGACTCGGAAATGGCAGCGCCCACAAAGAG
ACAATCGGAAAGCTGAGCAAAGTGAAGCTGTCCAGCCAGCTGAGCGTGTCCGACATCGATAAG
GCCAGCTCTGAGGCCCTTTGGTGCGCCCTGACAAGAGAACCTGGCTTCGACCCCAAAGAGGGA
CTGCCTGCCAATCCTGAGCGGCACATCAGAGTGAATGGCACCCATGTGTACGCCGGCGACAAC
ATCGGCCTGTTTCCAGTGTCTGCCGGATCTATCGCTCTGAGAGGCGGATATGCCGAGCTGGGC
AGCTCTTTCCATCACGCCAGGGTGTACAAGATCACAAGCGGCAAGAAACCCGCCTTTGCCATG
CTGAGAGTGTATACCATCGACCTGCTGCCTTACCGGAACCAGGACCTGTTCAGCGTGGAACTG
AAGCCCCAGACCATGAGCATGAGACAGGCCGAGAAGAAGCTGAGGGACGCCCTGGCTACAGGC
AACGCCGAATATCTTGGATGGCTGGTGGTGGATGACGAGCTGGTGGTCGATACCAGCAAGATC
GCCACCGACCAAGTGAAGGCTGTGGAAGCCGAACTGGGAACCATCAGACGTTGGCGCGTGGAC
GGCTTTTTCAGCCCCTCTAAGCTGAGACTGCGGCCCCTGCAGATGAGCAAAGAGGGCATCAAG
AAAGAGAGCGCCCCTGAGCTGTCCAAGATCATTGACAGACCTGGCTGGCTGCCCGCCGTGAAC
AAGCTGTTTTCTGACGGCAACGTGACCGTCGTGCGGAGAGATTCTCTGGGCAGAGTGCGCCTG
GAAAGCACAGCACATCTGCCCGTGACATGGAAGGTGCAGCCCAAGAAGAAGAGGAAGGTG

CdiCas9/NLS protein sequence (SEQ ID NO:30)

MKYHVGIDVGTFSVGLAAIEVDDAGMPIKTLSLVSHIHDSGLDPDKIKSAVTRLASSGIARRT
RRLYRRKRRRLQQLDKFIQRQGWPVIELEDYSDPLYPWKVRAELAASYIADEKERGEKLSVAL
RHIARHRGWRNPYAKVSSLYLPDEPSDAFKAIREEIKRASGQPVPETATVGQMVTLCELGTLK
LRGEGGVLSARLQQSDHAREIQEICRMQEIGQELYRKIIDVVFAAESPKGSASSRVGKDPLQP
GKNRALKASDAFQRYRIAALIGNLRVRVDGEKRILSVEEKNLVFDHLVNLAPKKEPEWVTIAE
ILGIDRGQLIGTATMTDDGERAGARPPTHDTNRSIVNSRIAPLVDWWKTASALEQHAMVKALS
NAEVDDFDSPEGAKVQAFFADLDDDVHAKLDSLHLPVGRAAYSEDTLVRLTRRMLADGVDLYT
ARLQEFGIEPSWTPPAPRIGEPVGNPAVDRVLKTVSRWLESATKTWGAPERVIIEHVREGFVT
EKRAREMDGDMRRRAARNAKLFQEMQEKLNVQGKPSRADLWRYQSVQRQNCQCAYCGSPITFS
NSEMDHIVPRAGQGSTNTRENLVAVCHRCNQSKGNTPFAIWAKNTSIEGVSVKEAVERTRHWV
TDTGMRSTDFKKFTKAVVERFQRATMDEEIDARSMESVAWMANELRSRVAQHFASHGTTVRVY
RGSLTAEARRASGISGKLEFLDGVGKSRLDRRHHAIDAAVIAFTSDYVAETLAVRSNLKQSQA
HRQEAPQWREFTGKDAEHRAAWRVWCQKMEKLSALLTEDLRDDRVVVMSNVRLRLGNGSAHEE
TIGKLSKVKLGSQLSVSDIDKASSEALWCALTREPDFDPKDGLPANPERHIRVNGTHVYAGDN
IGLFPVSAGSIALRGGYAELGSSFHHARVYKITSGKKPAFAMLRVYTIDLLPYRNQDLFSVEL
KPQTMSMRQAEKKLRDALATGNAEYLGWLVVDDELVVDTSKIATDQVKAVEAELGTIRRWRVD
GFFGDTRLRLRPLQMSKEGIKKESAPELSKIIDRPGWLPAVNKLFSEGNVTVVRRDSLGRVRL
ESTAHLPVTWKVQPKKKRKV

Bsm Cas9 sgRNA (SEQ ID NO: No 31)

NNNNNNNNNNNNNNNNNNNNGUCAUAGUUCCCCUAAGAUUAUUGAAACAAUGAUCUUAGGGUU
ACUAUGAUAAGGGCUUUCUACUUUAGGGGUAGAGAUGUCCCGCGGCGUUGGGGAUCGCCUAUU
GCCCUUAAAGGGCACUCCCCAUUUUAAUUUUUUU

Lrh Cas9 sgRNA (SEQ ID NO:32)
NNNNNNNNNNNNNNNNNNNNGUCUCAGGUAGAUGUCAGAUCAAUCAGAAAUGAUUGAUCUGAC
AUCUACGAGUUGAGAUCAAACAAAGCUUCAGCUGAGUUUCAAUUUCUGAGCCCAUGUUGGGCC
AUACAUAUGCCACCCGAGUGCAAAUCGGGUGGCUUUUUUU

Pex Cas9 sgRNA (SEQ ID NO:33)
NNNNNNNNNNNNNNNNNNNNGUUUCAGUAGUUGUUAGAAGAAUGAAAAUUCUUUUAACAACGA
AGUCGCCUUCGGGCGAGCUGAAAUCAAUUUGAUUAAAUAUUAGAUCCGGCUACUGAGGUCUUU
GACCUUAUCCGGAUUAACGAAGAGCCUCCGAGGAGGCUUUUUUU

Mca Cas9 sgRNA (SEQ ID NO:34)
NNNNNNNNNNNNNNNNNNNNGUUUUAGUGUUGUACAAUAUUUGGGUGAAAACCCAAAUAUUGU
ACAUCCUAAAUCAAGGCGCUUAAUUGCUGCCGUAAUUGCUGAAAGCGUAGCUUUCAGUUUUUU
U

Mga Cas9 sgRNA (SEQ ID NO:35)
NNNNNNNNNNNNNNNNNNNNGUUUUAGCACUGUACAAUACUUGUGUAAGCAAUAACGAAAAUU
AUUGCUUACACAAUUAUUGUCGUGCUAAAAUAAGGCGCUGUUAAUGCAGCUGCCGCAUCCGCC
AGAGCAUUUAUGCUCUGGCUUUUUUU

Agl Cas9 sgRNA (SEQ ID NO:36)
NNNNNNNNNNNNNNNNNNNNGUUUUGCCUUGAAUCCAAAGUAAGGCAUGGUAgaaaUAUUAUU
CCUGUGGAUUCAAGACAAAAUUUGAAAUGCAAACCGAUUCCCCGGCUGCAAGCCAGCCACACC
GGUCUUUCAAAGCAUUUUUUU

Amu Cas9 sgRNA (SEQ ID NO:37)
NNNNNNNNNNNNNNNNNNNNGUUUUGCCUUGAAUCCAAAACGGAUUCAAGACAAAAUUUGAAA
UGCAAACCGAUUUUCCUGACUGCCAGCCAGUCACACCGGUAACAAAAGCAUUUUUUU

Oki Cas9 sgRNA (SEQ ID NO:38)
NNNNNNNNNNNNNNNNNNNNGCUUCAGAUGUGUGUCAGAUCAAUGAgaaaUCAUUGAUCUGAC
ACACAGCAUUGAAGUAAAGCAAGAUUAAUUUCAAGCUUAAUUUUCUUCACAUUUUAUGUGCAG
AAGGGCUUAUGCCCACAAUACAUAAAAAGUCCGCAUUCACUUGCGGACUUUUAUUUUUUU

Bbo Cas9 sgRNA (SEQ ID NO:39)
NNNNNNNNNNNNNNNNNNNNGUUUCAAAUUCAAUCUAAAGCGAAAGCUAUACUUAUUAUUGAA
UUUGAAAUAAGGCUGUUCCUUCGUUAGUUCAGUCGAUUGCUCCUCCGGUAUUGCUUAUGCAUG
CCGGAGUUUUUU

Ace Cas9 sgRNA (SEQ ID NO:40)
NNNNNNNNNNNNNNNNNNNNGCUGGGGAGCCUGUCUGAAAAGACAGGCUACCUAGCAAGACCC
CUUCGUGGGGUCGCAUUCUUCACCCCCUCGCAGCAGCGAGGGGGUUCGUUUUUUU

Ahe Cas9 sgRNA (SEQ ID NO:41)
NNNNNNNNNNNNNNNNNNNNGUCAUAGUUCCCUCACAAGCCUCGAUGUGGAAACACAUCAAGG
CUUGCGAGGUUGCUAUGAUAAGGCAACAGGCCGCAAAGCACUGACCCGCAUUCCAAUGAAUGC
GGGUCAUCUACUUUUUUU

Wsu Cas9 sgRNA (SEQ ID NO:42)
NNNNNNNNNNNNNNNNNNNNGUUUCACAGGCUAAGCGGAUUUGCgaaaGCAAAUCCGUUCGAU
GCCUUGAAAUCAUCAAAAAGAUAUAAUAGACCCGCCCACUGUAUUGUACAUGGCGGGACUUUU
UUU

Nsa Cas9 sgRNA (SEQ ID NO:43)
NNNNNNNNNNNNNNNNNNNNGUUAUAAGACCCCUCAAAACCCCACCCUGUUACAAUGUUGUAA
CAGGGUAGGGUUAUUUGAGGGGUCUUAUAAUCAAGAACUGUUACAACAGUUCCAUUCUAGGGC
CCAUCUUCGGACGGGCCUCAGCCUUUUUUU

Rsy Cas9 sgRNA (SEQ ID NO:44)
NNNNNNNNNNNNNNNNNNNNGUUGUAGCCAGAGCGCAAUUCCCGAUCUGCUGAAAAGCAGAUC
GGGAAUUGCGCUUUGCUACUAACAAGCUGAAUCCGUUAGGAGUAAAUGCACCAAAUGAGAGGG
CCGGCUUAUGCCGGCCCUUUGCUUUUUUU

Cdi Cas9 sgRNA (SEQ ID NO:45)
NNNNNNNNNNNNNNNNNNNNACUGGGGUUCAGUUCUCAAAAACCCUGAUAGACUUCGAAAAGU
CACUAACUUAAUUAAAUAGAACUGAACCUCAGUAAGCAUUGGCUCGUUUCCAAUGUUGAUUGC
UCCGCCGGUGCUCCUUAUUAUUAAGGGCGCCGGCUUUCUUUUUUU

PexCas9-HN1HB1 fusion (SEQ ID NO:117)
MPKRKVSSAEGAAKEEPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQA
EVANQETKEDLPAENGETKTEESPASDEAGEKEAKSD*TGSG*MGKTHIIGVGLDLGGTYTGTFI
TSHPSDEAEHRDHSSAFTVVNSEKLSFSSKSRTAVRHRVRSYKGFDLRRRLLLLVAEYQLLQK
KQTLAPEERENLRIALSGYLKRRGYARTEAETDTSVLESLDPSVFSSAPSFTNFFNDSEPLNI
QWEAIANSPETTKALNKELSGQKEADFKKYIKTSFPEYSAKEILANYVEGRRAILDASKYIAN
LQSLGHKHRSKYLSDILQDMKRDSRITRLSEAFGSTDNLWRIIGNISNLQERAVRWYFNDAKF
EQGQEQLDAVKLKNVLVRALKYLRSDDKEWSASQKQIIQSLEQSGDVLDVLAGLDPDRTIPPY
EDQNNRRPPEDQTLYLNPKALSSEYGEKWKSWANKFAGAYPLLTEDLTEILKNTDRKSRIKIR
SDVLPDSDYRLAYILQRAFDRSIALDECSIRRTAEDFENGVVIKNEKLEDVLSGHQLEEFLEF
ANRYYQETAKAKNGLWFPENALLERADLHPPMKNKILNVIVGQALGVSPAEGTDFIEEIWNSK
VKGRSTVRSICNAIENERKTYGPYFSEDYKFVKTALKEGKTEKELSKKFAAVIKVLKMVSEVV
PFIGKELRLSDEAQSKFDNLYSLAQLYNLIETERNGFSKVSLAAHLENAWRMTMTDGSAQCCR
LPADCVRPFDGFIRKAIDRNSWEVAKRIAEEVKKSVDFTNGTVKIPVAIEANSFNFTASLTDL
KYIQLKEQKLKKKLEDIQRNEENQEKRWLSKEERIRADSHGICAYTGRPLDDVGEIDHIIPRS
LTLKKSESIYNSEVNLIFVSAQGNQEKKNNIYLLSNLAKNYLAAVFGTSDLSQITNEIESTVL
QLKAAGRLGYFDLLSEKERACARHALFLNSDSEARRAVIDVLGSRRKASVNGTQAWFVRSIFS
KVRQALAAWTQETGNELIFDAISVPAADSSEMRKRFAEYRPEFRKPKVQPVASHSIDAMCIYL
AACSDPFKTKRMGSQLAIYEPINFDNLFTGSCQVIQNTPRNFSDKTNIANSPIFKETIYAERF
LDIIVSRGEIFIGYPSNMPFEEKPNRISIGGKDPFSILSVLGAYLDKAPSSEKEKLTIYRVVK

NKAFELFSKVAGSKFTAEEDKAAKILEALHFVTVKQDVAATVSDLIKSKKELSKDSIENLAKQ
KGCLKKVEYSSKEFKFKGSLIIPAAVEWGKVLWNVFKENTAEELKDENALRKALEAAWPSSFG
TRNLHSKAKRVFSLPVVATQSGAVRIRRKTAFGDFVYQSQDTNNLYSSFPVKNGKLDWSSPII
HPALQNRNLTAYGYRFVDHDRSISMSEFREVYNKDDLMRIELAQGTSSRRYLRVEMPGEKFLA
WFGENSISLGSSFKFSVSEVFDNKIYTENAEFTKFLPKPREDNKHNGTIFFELVGPRVIFNYI
VGGAASSLKEIFSEAGKERS<u>PKKKRKV</u>*LEGGGGS***GKGDPKKPRGKMSSYAFFVQTCREEHKKK**
HPDASVNFSEFSKKCSERWKTMSAKEKGKFEDMAKADKARYEREMKTYIPPKGE

PexCas9-HN1H1G fusion (SEQ ID NO:118)

MPKRKVSSAEGAAKEEPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQA
EVANQETKEDLPAENGETKTEESPASDEAGEKEAKSD*TGSG*MGKTHIIGVGLDLGGTYTGTFI
TSHPSDEAEHRDHSSAFTVVNSEKLSFSSKSRTAVRHRVRSYKGFDLRRRLLLLVAEYQLLQK
KQTLAPEERENLRIALSGYLKRRGYARTEAETDTSVLESLDPSVFSSAPSFTNFFNDSEPLNI
QWEAIANSPETTKALNKELSGQKEADFKKYIKTSFPEYSAKEILANYVEGRRAILDASKYIAN
LQSLGHKHRSKYLSDILQDMKRDSRITRLSEAFGSTDNLWRIIGNISNLQERAVRWYFNDAKF
EQGQEQLDAVKLKNVLVRALKYLRSDDKEWSASQKQIIQSLEQSGDVLDVLAGLDPDRTIPPY
EDQNNRRPPEDQTLYLNPKALSSEYGEKWKSWANKFAGAYPLLTEDLTEILKNTDRKSRIKIR
SDVLPDSDYRLAYILQRAFDRSIALDECSIRRTAEDFENGVVIKNEKLEDVLSGHQLEEFLEF
ANRYYQETAKAKNGLWFPENALLERADLHPPMKNKILNVIVGQALGVSPAEGTDFIEEIWNSK
VKGRSTVRSICNAIENERKTYGPYFSEDYKFVKTALKEGKTEKELSKKFAAVIKVLKMVSEVV
PFIGKELRLSDEAQSKFDNLYSLAQLYNLIETERNGFSKVSLAAHLENAWRMTMTDGSAQCCR
LPADCVRPFDGFIRKAIDRNSWEVAKRIAEEVKKSVDFTNGTVKIPVAIEANSFNFTASLTDL
KYIQLKEQKLKKKLEDIQRNEENQEKRWLSKEERIRADSHGICAYTGRPLDDVGEIDHIIPRS
LTLKKSESIYNSEVNLIFVSAQGNQEKKNNIYLLSNLAKNYLAAVFGTSDLSQITNEIESTVL
QLKAAGRLGYFDLLSEKERACARHALFLNSDSEARRAVIDVLGSRRKASVNGTQAWFVRSIFS
KVRQALAAWTQETGNELIFDAISVPAADSSEMRKRFAEYRPEFRKPKVQPVASHSIDAMCIYL
AACSDPFKTKRMGSQLAIYEPINFDNLFTGSCQVIQNTPRNFSDKTNIANSPIFKETIYAERF
LDIIVSRGEIFIGYPSNMPFEEKPNRISIGGKDPFSILSVLGAYLDKAPSSEKEKLTIYRVVK
NKAFELFSKVAGSKFTAEEDKAAKILEALHFVTVKQDVAATVSDLIKSKKELSKDSIENLAKQ
KGCLKKVEYSSKEFKFKGSLIIPAAVEWGKVLWNVFKENTAEELKDENALRKALEAAWPSSFG
TRNLHSKAKRVFSLPVVATQSGAVRIRRKTAFGDFVYQSQDTNNLYSSFPVKNGKLDWSSPII
HPALQNRNLTAYGYRFVDHDRSISMSEFREVYNKDDLMRIELAQGTSSRRYLRVEMPGEKFLA
WFGENSISLGSSFKFSVSEVFDNKIYTENAEFTKFLPKPREDNKHNGTIFFELVGPRVIFNYI
VGGAASSLKEIFSEAGKERS<u>PKKKRKV</u>*LEGGGGS***STDHPKYSDMIVAAIQAEKNRAGSSRQSI**
QKYIKSHYKVGENADSQIKLSIKRLVTTGVLKQTKGVGASGSFRLAKSDEP

BsmCas9-HN1HB1 fusion (SEQ ID NO:119)

MPKRKVSSAEGAAKEEPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQA
EVANQETKEDLPAENGETKTEESPASDEAGEKEAKSD*TGSG*MNYKMGLDIGIASVGWAVINLD
LKRIEDLGVRIFDKAEHPQNGESLALPRRIARSARRRLRRRKHRLERIRRLLVSENVLTKEEM
NLLFKQKKQIDVWQLRVDALERKLNNDELARVLLHLAKRRGFKSNRKSERNSKESSEFLKNIE
ENQSILAQYRSVGEMIVKDSKFAYHKRNKLDSYSNMIARDDLEREIKLIFEKQREFNNPVCTE
RLEEKYLNIWSSQRPFASKEDIEKKVGFCTFEPKEKRAPKATYTFQSFIVWEHINKLRLVSPD
ETRALTEIERNLLYKQAFSKNKMTYYDIRKLLNLSDDIHFKGLLYDPKSSLKQIENIRFLELD
SYHKIRKCIENVYGKDGIRMFNETDIDTFGYALTIFKDDEDIVAYLQNEYITKNGKRVSNLAN
KVYDKSLIDELLNLSFSKFAHLSMKAIRNILPYMEQGEIYSKACELAGYNFTGPKKKEKALLL

PVIPNIANPVVMRALTQSRKVVNAIIKKYGSPVSIHIELARDLSHSFDERKKIQKDQTENRKK
NETAIKQLIEYELTKNPTGLDIVKFKLWSEQQGRCMYSLKPIELERLLEPGYVEVDHILPYSR
SLDDSYANKVLVLTKENREKGNHTPVEYLGLGSERWKKFEKFVLANKQFSKKKKQNLLRLRYE
ETEEKEFKERNLNDTRYISKFFANFIKEHLKFADGDGGQKVYTINGKITAHLRSRWDFNKNRE
ESDLHHAVDAVIVACATQGMIKKITEFYKAREQNKESAKKKEPIFPQPWPHFADELKARLSKF
PQESIEAFALGNYDRKKLESLRPVFVSRMPKRSVTGAAHQETLRRCVGIDEQSGKIQTAVKTK
LSDIKLDKDGHFPMYQKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGEPGPVIRTVK
IIDTKNKVVHLDGSKTVAYNSNIVRTDVFEKDGKYYCVPVYTMDIMKGTLPNKAIEANKPYSE
WKEMTEEYTFQFSLFPNDLVRIVLPREKTIKTSTNEEIIIKDIFAYYKTIDSATGGLELISHD
RNFSLRGVGSKTLKRFEKYQVDVLGNIHKVKGEKRVGLAAPTNQKKGKTVDSLQSVSDPKKKR
KV*LEGGGGS***GKGDPKKPRGKMSSYAFFVQTCREEHKKKHPDASVNFSEFSKKCSERWKTMSAK**
EKGKFEDMAKADKARYEREMKTYIPPKGE

BsmCas9-HN1H1G fusion (SEQ ID NO:120)

MPKRKVSSAEGAAKEEPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQA
EVANQETKEDLPAENGETKTEESPASDEAGEKEAKSD*TGSG*MNYKMGLDIGIASVGWAVINLD
LKRIEDLGVRIFDKAEHPQNGESLALPRRIARSARRRLRRRKHRLERIRRLLVSENVLTKEEM
NLLFKQKKQIDVWQLRVDALERKLNNDELARVLLHLAKRRGFKSNRKSERNSKESSEFLKNIE
ENQSILAQYRSVGEMIVKDSKFAYHKRNKLDSYSNMIARDDLEREIKLIFEKQREFNNPVCTE
RLEEKYLNIWSSQRPFASKEDIEKKVGFCTFEPKEKRAPKATYTFQSFIVWEHINKLRLVSPD
ETRALTEIERNLLYKQAFSKNKMTYYDIRKLLNLSDDIHFKGLLYDPKSSLKQIENIRFLELD
SYHKIRKCIENVYGKDGIRMFNETDIDTFGYALTIFKDDEDIVAYLQNEYITKNGKRVSNLAN
KVYDKSLIDELLNLSFSKFAHLSMKAIRNILPYMEQGEIYSKACELAGYNFTGPKKKEKALLL
PVIPNIANPVVMRALTQSRKVVNAIIKKYGSPVSIHIELARDLSHSFDERKKIQKDQTENRKK
NETAIKQLIEYELTKNPTGLDIVKFKLWSEQQGRCMYSLKPIELERLLEPGYVEVDHILPYSR
SLDDSYANKVLVLTKENREKGNHTPVEYLGLGSERWKKFEKFVLANKQFSKKKKQNLLRLRYE
ETEEKEFKERNLNDTRYISKFFANFIKEHLKFADGDGGQKVYTINGKITAHLRSRWDFNKNRE
ESDLHHAVDAVIVACATQGMIKKITEFYKAREQNKESAKKKEPIFPQPWPHFADELKARLSKF
PQESIEAFALGNYDRKKLESLRPVFVSRMPKRSVTGAAHQETLRRCVGIDEQSGKIQTAVKTK
LSDIKLDKDGHFPMYQKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGEPGPVIRTVK
IIDTKNKVVHLDGSKTVAYNSNIVRTDVFEKDGKYYCVPVYTMDIMKGTLPNKAIEANKPYSE
WKEMTEEYTFQFSLFPNDLVRIVLPREKTIKTSTNEEIIIKDIFAYYKTIDSATGGLELISHD
RNFSLRGVGSKTLKRFEKYQVDVLGNIHKVKGEKRVGLAAPTNQKKGKTVDSLQSVSDPKKKR
KV*LEGGGGS***STDHPKYSDMIVAAIQAEKNRAGSSRQSIQKYIKSHYKVGENADSQIKLSIKRL**
VTTGVLKQTKGVGASGSFRLAKSDEP

LrhCas9-HN1HB1 fusion (SEQ ID NO:121)

MPKRKVSSAEGAAKEEPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQA
EVANQETKEDLPAENGETKTEESPASDEAGEKEAKSD*TGSG*MTKLNQPYGIGLDIGSNSIGFA
VVDANSHLLRLKGETAIGARLFREGQSAADRRGSRTTRRRLSRTRWRLSFLRDFFAPHITKID
PDFFLRQKYSEISPKDKDRFKYEKRLFNDRTDAEFYEDYPSMYHLRLHLMTHTHKADPREIFL
AIHHILKSRGHFLTPGAAKDFNTDKVDLEDIFPALTEAYAQVYPDLELTFDLAKADDFKAKLL
DEQATPSDTQKALVNLLLSSDGEKEIVKKRKQVLTEFAKAITGLKTKFNLALGTEVDEADASN
WQFSMGQLDDKWSNIETSMTDQGTEIFEQIQELYRARLLNGIVPAGMSLSQAKVADYGQHKED
LELFKTYLKKLNDHELAKTIRGLYDRYINGDDAKPFLREDFVKALTKEVTAHPNEVSEQLLNR
MGQANFMLKQRTKANGAIPIQLQQRELDQIIANQSKYYDWLAAPNPVEAHRWKMPYQLDELLN

FHIPYYVGPLITPKQQAESGENVFAWMVRKDPSGNITPYNFDEKVDREASANTFIQRMKTTDT
YLIGEDVLPKQSLLYQKYEVLNELNNVRINNECLGTDQKQRLIREVFERHSSVTIKQVADNLV
AHGDFARRPEIRGLADEKRFLSSLSTYHQLKEILHEAIDDPTKLLDIENIITWSTVFEDHTIF
ETKLAEIEWLDPKKINELSGIRYRGWGQFSRKLLDGLKLGNGHTVIQELMLSNHNLMQILADE
TLKETMTELNQDKLKTDDIEDVINDAYTSPSNKKALRQVLRVVEDIKHAANGQDPSWLFIETA
DGTGTAGKRTQSRQKQIQTVYANAAQELIDSAVRGELEDKIADKASFTDRLVLYFMQGGRDIY
TGAPLNIDQLSHYDIDHILPQSLIKDDSLDNRVLVNATINREKNNVFASTLFAGKMKATWRKW
HEAGLISGRKLRNLMLRPDEIDKFAKGFVARQLVETRQIIKLTEQIAAAQYPNTKIIAVKAGL
SHQLREELDFPKNRDVNHYHHAFDAFLAARIGTYLLKRYPKLAPFFTYGEFAKVDVKKFREFN
FIGALTHAKKNIIAKDTGEIVWDKERDIRELDRIYNFKRMLITHEVYFETADLFKQTIYAAKD
SKERGGSKQLIPKKQGYPTQVYGGYTQESGSYNALVRVAEADTTAYQVIKISAQNASKIASAN
LKSREKGKQLLNEIVVKQLAKRRKNWKPSANSFKIVIPRFGMGTLFQNAKYGLFMVNSDTYYR
NYQELWLSRENQKLLKKLFSIKYEKTQMNHDALQVYKAIIDQVEKFFKLYDINQFRAKLSDAI
ERFEKLPINTDGNKIGKTETLRQILIGLQANGTRSNVKNLGIKTDLGLLQVGSGIKLDKDTQI
VYQSPSGLFKRRIPLADLPKKKRKV*LEGGGGS***GKGDPKKPRGKMSSYAFFVQTCREEHKKKHP**
DASVNFSEFSKKCSERWKTMSAKEKGKFEDMAKADKARYEREMKTYIPPKGE

LrhCas9-HN1H1G fusion (SEQ ID NO:122)

MPKRKVSSAEGAAKEEPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQA
EVANQETKEDLPAENGETKTEESPASDEAGEKEAKSD*TGSG*MTKLNQPYGIGLDIGSNSIGFA
VVDANSHLLRLKGETAIGARLFREGQSAADRRGSRTTRRRLSRTRWRLSFLRDFFAPHITKID
PDFFLRQKYSEISPKDKDRFKYEKRLFNDRTDAEFYEDYPSMYHLRLHLMTHTHKADPREIFL
AIHHILKSRGHFLTPGAAKDFNTDKVDLEDIFPALTEAYAQVYPDLELTFDLAKADDFKAKLL
DEQATPSDTQKALVNLLLSSDGEKEIVKKRKQVLTEFAKAITGLKTKFNLALGTEVDEADASN
WQFSMGQLDDKWSNIETSMTDQGTEIFEQIQELYRARLLNGIVPAGMSLSQAKVADYGQHKED
LELFKTYLKKLNDHELAKTIRGLYDRYINGDDAKPFLREDFVKALTKEVTAHPNEVSEQLLNR
MGQANFMLKQRTKANGAIPIQLQQRELDQIIANQSKYYDWLAAPNPVEAHRWKMPYQLDELLN
FHIPYYVGPLITPKQQAESGENVFAWMVRKDPSGNITPYNFDEKVDREASANTFIQRMKTTDT
YLIGEDVLPKQSLLYQKYEVLNELNNVRINNECLGTDQKQRLIREVFERHSSVTIKQVADNLV
AHGDFARRPEIRGLADEKRFLSSLSTYHQLKEILHEAIDDPTKLLDIENIITWSTVFEDHTIF
ETKLAEIEWLDPKKINELSGIRYRGWGQFSRKLLDGLKLGNGHTVIQELMLSNHNLMQILADE
TLKETMTELNQDKLKTDDIEDVINDAYTSPSNKKALRQVLRVVEDIKHAANGQDPSWLFIETA
DGTGTAGKRTQSRQKQIQTVYANAAQELIDSAVRGELEDKIADKASFTDRLVLYFMQGGRDIY
TGAPLNIDQLSHYDIDHILPQSLIKDDSLDNRVLVNATINREKNNVFASTLFAGKMKATWRKW
HEAGLISGRKLRNLMLRPDEIDKFAKGFVARQLVETRQIIKLTEQIAAAQYPNTKIIAVKAGL
SHQLREELDFPKNRDVNHYHHAFDAFLAARIGTYLLKRYPKLAPFFTYGEFAKVDVKKFREFN
FIGALTHAKKNIIAKDTGEIVWDKERDIRELDRIYNFKRMLITHEVYFETADLFKQTIYAAKD
SKERGGSKQLIPKKQGYPTQVYGGYTQESGSYNALVRVAEADTTAYQVIKISAQNASKIASAN
LKSREKGKQLLNEIVVKQLAKRRKNWKPSANSFKIVIPRFGMGTLFQNAKYGLFMVNSDTYYR
NYQELWLSRENQKLLKKLFSIKYEKTQMNHDALQVYKAIIDQVEKFFKLYDINQFRAKLSDAI
ERFEKLPINTDGNKIGKTETLRQILIGLQANGTRSNVKNLGIKTDLGLLQVGSGIKLDKDTQI
VYQSPSGLFKRRIPLADLPKKKRKV*LEGGGGS***STDHPKYSDMIVAAIQAEKNRAGSSRQSIQK**
YIKSHYKVGENADSQIKLSIKRLVTTGVLKQTKGVGASGSFRLAKSDEP

McaCas9-HN1HB1 fusion (SEQ ID NO:123)

MPKRKVSSAEGAAKEEPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQA
EVANQETKEDLPAENGETKTEESPASDEAGEKEAKSD*TGSG*MEKKRKVTLGFDLGIASVGWAI
VDSETNQVYKLGSRLFDAPDTNLERRTQRGTRRLLRRRKYRNQKFYNLVKRTEVFGLSSREAI
ENRFRELSIKYPNIIELKTKALSQEVCPDEIAWILHDYLKNRGYFYDEKETKEDFDQQTVESM
PSYKLNEFYKKYGYFKGALSQPTESEMKDNKDLKEAFFFDFSNKEWLKEINYFFNVQKNILSE
TFIEEFKKIFSFTRDISKGPGSDNMPSPYGIFGEFGDNGQGGRYEHIWDKNIGKCSIFTNEQR
APKYLPSALIFNFLNELANIRLYSTDKKNIQPLWKLSSIDKLNILLNLFNLPISEKKKKLTST
NINDIVKKESIKSIMLSVEDIDMIKDEWAGKEPNVYGVGLSGLNIEESAKENKFKFQDLKILN
VLINLLDNVGIKFEFKDRSDIIKNLELLDNLYLFLIYQKESNNKDSSIDLFIAKNKSLNIENL
KLKLKEFLLGAGNEFENHNSKTHSLSKKAIDAILPKLLDNNEGWNLEAIKNYDEEIKSQIEDN
SSLMAKQDKKYLNDNFLKDAILPPNVKVTFQQAILIFNKIIQKFSKDFEIDKVVIELAREMTQ
DQENDALKGIAKAQKSKKSLVEERLEANNIDKSVFNDKYEKLIYKIFLWISQDFKDPYTGAKI
SANEIVDNKVEIDHIIPYSLCFDDSSANKVLVHKQSNQEKSNSLPYEYIKQGHSGWNWDEFTK
YVKRVFVNNVDSILSKKERLKKSENLLTTSYDGYEKLGFLARNLNDTRYATILFRDQLNNYAE
HHLIDNKKMFKVIAMNGAVTSFIRKNMSYDNKLRLKDRSDFSHHAYDAAIIALFSNKTKTLYN
LIDPSLNGIISKRSEGYWVIEDRYTGEIKELKKEDWTSIKNNVQARKIAKEIEEYLIDLDDEV
FFSRKTKRKTNRQLYNETIYGIAAKTDEDGITNYYKKEKFSILDDKDIYLRLLREREKFVINQ
SNPEVIDQIIEIIESYGKENNIPSRDEAINIKYTKNKINYNLYLKQYMRSLTKSLDQFSEGFI
NQMIANKTFVLYNPTKNTTRKIKFLRLVNDVKINDIRKNQVINKFNGKNNEPKAFYENINSLG
AIVFKSSANNFKTLSINTQIAIFGDKNWDIEDFKTYNMEKIEKYKEIYGIDKTYNFHSFIFPG
TILLDKQNKEFYYISSIQTVNDQIELKFLNKIEFKNDDNTSGANKPPRRLRFGIKSIMNNYEQ
VDISPFGINKKIFE<u>PKKKRKV</u>*LEGGGGS***GKGDPKKPRGKMSSYAFFVQTCREEHKKKHPDASV**
NFSEFSKKCSERWKTMSAKEKGKFEDMAKADKARYEREMKTYIPPKGE

McaCas9-HN1H1G fusion (SEQ ID NO:124)
MPKRKVSSAEGAAKEEPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQA
EVANQETKEDLPAENGETKTEESPASDEAGEKEAKSD*TGSG*MEKKRKVTLGFDLGIASVGWAI
VDSETNQVYKLGSRLFDAPDTNLERRTQRGTRRLLRRRKYRNQKFYNLVKRTEVFGLSSREAI
ENRFRELSIKYPNIIELKTKALSQEVCPDEIAWILHDYLKNRGYFYDEKETKEDFDQQTVESM
PSYKLNEFYKKYGYFKGALSQPTESEMKDNKDLKEAFFFDFSNKEWLKEINYFFNVQKNILSE
TFIEEFKKIFSFTRDISKGPGSDNMPSPYGIFGEFGDNGQGGRYEHIWDKNIGKCSIFTNEQR
APKYLPSALIFNFLNELANIRLYSTDKKNIQPLWKLSSIDKLNILLNLFNLPISEKKKKLTST
NINDIVKKESIKSIMLSVEDIDMIKDEWAGKEPNVYGVGLSGLNIEESAKENKFKFQDLKILN
VLINLLDNVGIKFEFKDRSDIIKNLELLDNLYLFLIYQKESNNKDSSIDLFIAKNKSLNIENL
KLKLKEFLLGAGNEFENHNSKTHSLSKKAIDAILPKLLDNNEGWNLEAIKNYDEEIKSQIEDN
SSLMAKQDKKYLNDNFLKDAILPPNVKVTFQQAILIFNKIIQKFSKDFEIDKVVIELAREMTQ
DQENDALKGIAKAQKSKKSLVEERLEANNIDKSVFNDKYEKLIYKIFLWISQDFKDPYTGAKI
SANEIVDNKVEIDHIIPYSLCFDDSSANKVLVHKQSNQEKSNSLPYEYIKQGHSGWNWDEFTK
YVKRVFVNNVDSILSKKERLKKSENLLTTSYDGYEKLGFLARNLNDTRYATILFRDQLNNYAE
HHLIDNKKMFKVIAMNGAVTSFIRKNMSYDNKLRLKDRSDFSHHAYDAAIIALFSNKTKTLYN
LIDPSLNGIISKRSEGYWVIEDRYTGEIKELKKEDWTSIKNNVQARKIAKEIEEYLIDLDDEV
FFSRKTKRKTNRQLYNETIYGIAAKTDEDGITNYYKKEKFSILDDKDIYLRLLREREKFVINQ
SNPEVIDQIIEIIESYGKENNIPSRDEAINIKYTKNKINYNLYLKQYMRSLTKSLDQFSEGFI
NQMIANKTFVLYNPTKNTTRKIKFLRLVNDVKINDIRKNQVINKFNGKNNEPKAFYENINSLG
AIVFKSSANNFKTLSINTQIAIFGDKNWDIEDFKTYNMEKIEKYKEIYGIDKTYNFHSFIFPG
TILLDKQNKEFYYISSIQTVNDQIELKFLNKIEFKNDDNTSGANKPPRRLRFGIKSIMNNYEQ

VDISPFGINKKIFEPKKKRKV*LEGGGGS***STDHPKYSDMIVAAIQAEKNRAGSSRQSIQKYIKS**
HYKVGENADSQIKLSIKRLVTTGVLKQTKGVGASGSFRLAKSDEP

CLAIMS

1. A system for modifying a chromosomal sequence in a eukaryotic cell, the system comprising: (a) an engineered *Bacillus smithii* Cas9 protein; and (b) an engineered guide RNA, wherein the engineered guide RNA is designed to complex with the engineered Cas9 protein and the engineered guide RNA comprises a 5’ guide sequence designed to hybridize with a target sequence in a double-stranded sequence, wherein the target sequence is 5’ to a protospacer adjacent motif (PAM) and the PAM comprises the sequence 5'-NNNNCAAA-3', wherein N is A, C, G, or T.

2. The system of claim 1, wherein the engineered Cas9 protein comprises at least one modification relative to its wild-type counterpart.

3. The system of claim 2, wherein the at least one modification comprises:

   (a) the addition of at least one heterologous domain; or

   (b) a substitution of one or more amino acids, an insertion of one or more amino acids, a deletion of one or more amino acids, or combination thereof, wherein the at least one modification is within a RuvC domain, HNH domain, REC domain, PAM interacting domain, or combination thereof.

4. The system of claim 3, wherein the at least one heterologous domain is a nuclear localization signal, a cell-penetrating domain, a marker domain, a chromatin modulating motif, an epigenetic modification domain, a transcriptional regulation domain, an RNA aptamer binding domain, or combination thereof.

5. The system of any one of claims 1 to 4, wherein the engineered Cas9 protein is a nuclease and cleaves both strands of a double-stranded sequence, is a nickase and cleaves one strand of a double-stranded sequence, or has no nuclease or nickase activity.

6. The system of any one of claims 1 to 5, wherein the engineered guide RNA is a single molecule.

7. The system of any one of claims 1 to 6, wherein the engineered guide RNA sequence is optimized to facilitate base-paring within the engineered guide RNA, minimize base-paring within the engineered guide RNA, increase stability of the engineered guide RNA, facilitate transcription of the engineered guide RNA in a eukaryotic cell, or a combination thereof.

8. The system of any one of claims 1 to 7, wherein the engineered guide RNA comprises the nucleotide sequence of SEQ ID NO:31.

9. The system of any one of claims 1 to 8, wherein the engineered Cas9 protein has an amino acid sequence having at least about 90% sequence identity to SEQ ID NO:2, 119, or 120.

10. The system of any one of claims 1 to 9, wherein the engineered Cas9 protein has an amino acid sequence as set forth in SEQ ID NO:2, 119, or 120.

11. A plurality of nucleic acids encoding the system of any one of claims 1 to 10, the plurality of nucleic acid comprising at least one nucleic acid encoding the engineered Cas9 protein, and at least one nucleic acid encoding the engineered guide RNA.

12. The plurality of nucleic acids of claim 11, wherein the at least one nucleic acid encoding the engineered Cas9 protein is RNA.

13. The plurality of nucleic acids of claim 11, wherein the at least one nucleic acid encoding the engineered Cas9 protein is DNA.

14. The plurality of nucleic acids of any one of claims 11 to 13, wherein the at least one nucleic acid encoding the engineered Cas9 protein is codon optimized for expression in a eukaryotic cell.

15. The plurality of nucleic acids of claim 14, wherein the eukaryotic cell is a human cell, a non-human mammalian cell, a non-mammalian vertebrate cell, an invertebrate cell, a plant cell, or a single cell eukaryotic organism.

16. The plurality of nucleic acids of claim 11, wherein the at least one nucleic acid encoding the engineered guide RNA is DNA.

17. The plurality of nucleic acids of any one of claims 11 to 16, wherein the at least one nucleic acid encoding the engineered Cas9 protein is operably linked to a phage promoter sequence for *in vitro* RNA synthesis or protein expression in a bacterial cell, and the at least one nucleic acid encoding the engineered guide RNA is operably linked to a phage promoter sequence for *in vitro* RNA synthesis.

18. The plurality of nucleic acids of any one of claims 11 to 16, wherein the at least one nucleic acid encoding the engineered Cas9 protein is operably linked to a eukaryotic promoter sequence for expression in a eukaryotic cell, and the at least one nucleic acid encoding the engineered guide RNA is operably linked to a eukaryotic promoter sequence for expression in a eukaryotic cell.

19. At least one vector comprising the plurality of nucleic acids of any one of claims 11 to 18, optionally wherein the vector is selected from: a plasmid vector, a viral vector, or a self-replicating viral RNA replicon.

20. A eukaryotic cell comprising at least one system comprising an engineered Cas9 protein and an engineered guide RNA as defined in claims 1 to 10, at least one nucleic acid as defined in claims 11 to 18, or at least one vector as defined in claim 19, optionally wherein the eukaryotic cell is selected from: a human cell, a non-human mammalian cell, a plant cell, a non-mammalian vertebrate cell, an invertebrate cell, or a single cell eukaryotic organism, optionally wherein the eukaryotic cell is *in vivo*, *ex vivo*, or *in vitro*.