

(51) International Patent Classification:  
C12N 15/85 (2006.01)(21) International Application Number:  
PCT/US2014/027335(22) International Filing Date:  
14 March 2014 (14.03.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

|            |                               |    |
|------------|-------------------------------|----|
| 61/799,647 | 15 March 2013 (15.03.2013)    | US |
| 61/838,178 | 21 June 2013 (21.06.2013)     | US |
| 61/838,148 | 21 June 2013 (21.06.2013)     | US |
| 61/921,007 | 26 December 2013 (26.12.2013) | US |

(71) Applicant: THE GENERAL HOSPITAL CORPORATION [US/US]; 55 Fruit Street, Boston, Massachusetts 02114 (US).

(72) Inventors: JOUNG, J. Keith; 1 Magnolia Way, Winchester, Massachusetts 01890 (US). MAEDER, Morgan; 1569 Beacon Street, Apt. 32, Brookline, Massachusetts 02446 (US).

(74) Agents: KUGLER DEYOUNG, Janice et al.; Fish &amp; Richardson P.C., P.O. Box 1022, Minneapolis, Minnesota 55440-1022 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

**Published:**

— without international search report and to be republished upon receipt of that report (Rule 48.2(g))

(54) Title: RNA-GUIDED TARGETING OF GENETIC AND EPIGENOMIC REGULATORY PROTEINS TO SPECIFIC GENOMIC LOCI

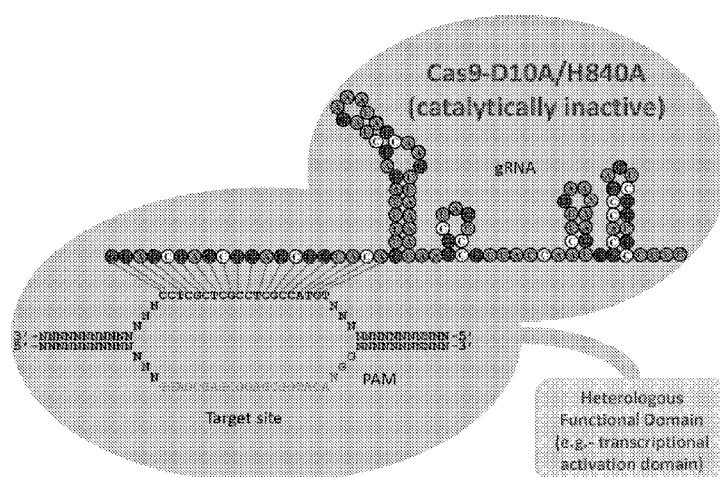


FIG. 1C

(57) Abstract: Methods and constructs for RNA-guided targeting of heterologous functional domains such as transcriptional activators to specific genomic loci.

# **RNA-GUIDED TARGETING OF GENETIC AND EPIGENOMIC REGULATORY PROTEINS TO SPECIFIC GENOMIC LOCI**

## **CLAIM OF PRIORITY**

5           This application claims the benefit of U.S. Patent Application Serial Nos. 61/799,647, filed on March 15, 2013; 61/838,178, filed on June 21, 2013; 61/838,148, filed on June 21, 2013; and 61/921,007, filed on December 26, 2013. The entire contents of the foregoing are hereby incorporated by reference.

## **FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT**

10           This invention was made with Government support under Grant No. DP1GM105378 awarded by the National Institutes of Health and W911NF-11-2-0056 awarded by the Defense Advanced Research Projects Agency (DARPA) of the Department of Defense. The Government has certain rights in the invention.

## **TECHNICAL FIELD**

15           This invention relates to methods and constructs for RNA-guided targeting of genetic and epigenomic regulatory proteins, e.g., transcriptional activators, histone modification enzymes, DNA methylation modifiers, to specific genomic loci.

## **BACKGROUND**

          Clustered Regulatory Interspaced Short Palindromic Repeats (CRISPR), and  
20   CRISPR-associated (cas) genes, referred to as CRISPR/Cas systems, are used by various bacteria and archaea to mediate defense against viruses and other foreign nucleic acid. These systems use small RNAs to detect and silence foreign nucleic acids in a sequence-specific manner.

          Three types of CRISPR/Cas systems have been described (Makarova et al., Nat.  
25   Rev. Microbiol. 9, 467 (2011); Makarova et al., Biol. Direct 1, 7 (2006); Makarova et al., Biol. Direct 6, 38 (2011)). Recent work has shown that Type II CRISPR/Cas systems can be engineered to direct targeted double-stranded DNA breaks in vitro to specific sequences by using a single “guide RNA” with complementarity to the DNA target site and a Cas9 nuclease (Jinek et al., Science 2012; 337:816–821). This targetable Cas9-based system also  
30   works in cultured human cells (Mali et al., Science. 2013 Feb 15;339(6121):823-6; Cong et al., Science. 2013 Feb 15;339(6121):819-23) and in vivo in zebrafish (Hwang and Fu et al.,

Nat Biotechnol. 2013 Mar;31(3):227-9) for inducing targeted alterations into endogenous genes.

### SUMMARY

At least in part, the present invention is based on the development of a fusion protein including a heterologous functional domain (e.g., a transcriptional activation domain) fused to a Cas9 nuclease that has had its nuclease activity inactivated by mutations (also known as “dCas9”). While published studies have used guide RNAs to target catalytically active and inactive Cas9 nuclease proteins to specific genomic loci, no work has yet adapted the use of this system to recruit additional effector domains. This work also provides the first demonstration of an RNA-guided process that results in an increase (rather than a decrease) in the level of expression of a target gene.

In addition, the present disclosure provides the first demonstration that multiplex gRNAs can be used to bring multiple dCas9-VP64 fusions to a single promoter, thereby resulting in synergistic activation of transcription.

Thus, in a first aspect, the invention provides fusion proteins comprising a catalytically inactive CRISPR associated 9 (dCas9) protein linked to a heterologous functional domain (HFD) that modifies gene expression, histones, or DNA, e.g., transcriptional activation domain, transcriptional repressors (e.g., silencers such as Heterochromatin Protein 1 (HP1), e.g., HP1 $\alpha$  or HP1 $\beta$ , or a transcriptional repression domain, e.g., Krueppel-associated box (KRAB) domain, ERF repressor domain (ERD), or mSin3A interaction domain (SID)), enzymes that modify the methylation state of DNA (e.g., DNA methyltransferase (DNMT) or Ten-Eleven Translocation (TET) proteins, e.g., TET1, also known as Tet Methylcytosine Dioxygenase 1), or enzymes that modify histone subunit (e.g., histone acetyltransferases (HAT), histone deacetylases (HDAC), or histone demethylases). In some embodiments, the heterologous functional domain is a transcriptional activation domain, e.g., a transcriptional activation domain from VP64 or NF- $\kappa$ B p65; an enzyme that catalyzes DNA demethylation, e.g., a TET; or histone modification (e.g., LSD1, histone methyltransferase, HDACs, or HATs) or a transcription silencing domain, e.g., from Heterochromatin Protein 1 (HP1), e.g., HP1 $\alpha$  or HP1 $\beta$ ; or a biological tether, e.g., CRISPR/Cas Subtype Ypest protein 4 (Csy4), MS2, or lambda N protein.

In some embodiments, the catalytically inactive Cas9 protein is from *S. pyogenes*.

In some embodiments, the catalytically inactive Cas9 protein comprises mutations at D10, E762, H983, or D986; and at H840 or N863, e.g., at D10 and H840, e.g., D10A or D10N and H840A or H840N or H840Y.

In some embodiments, the heterologous functional domain is linked to the N terminus or C terminus of the catalytically inactive Cas9 protein, with an optional intervening linker, wherein the linker does not interfere with activity of the fusion protein.

In some embodiments, the fusion protein includes one or both of a nuclear localization sequence and one or more epitope tags, e.g., c-myc, 6His, or FLAG tags, on the N-terminus, C-terminus, or in between the catalytically inactive CRISPR associated 9 (Cas9) protein and the heterologous functional domain, optionally with one or more intervening linkers.

In further aspect, the invention provides nucleic acids encoding the fusion proteins described herein, as well as expression vectors including the nucleic acids, and host cells expressing the fusion proteins.

In an additional aspect, the invention provides methods for increasing expression of a target gene in a cell. The methods include expressing a Cas9-HFD fusion protein as described herein in the cell, e.g., by contacting the cell with an expression vector including a sequence encoding the fusion protein, and also expressing in the cell one or more guide RNAs with complementarity directed to the target gene, e.g., by contacting the cell with one or more expression vectors comprising nucleic acid sequences encoding one or more guide RNAs.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Methods and materials are described herein for use in the present invention; other, suitable methods and materials known in the art can also be used. The materials, methods, and examples are illustrative only and not intended to be limiting. All publications, patent applications, patents, sequences, database entries, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control.

Other features and advantages of the invention will be apparent from the following detailed description and figures, and from the claims.

## DESCRIPTION OF DRAWINGS

The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

FIG. 1A is a schematic illustration showing a single guide RNA (sgRNA) recruiting Cas9 nuclease to a specific DNA sequence and thereby introducing targeted alterations. The sequence of the guide RNA shown is  
GGAGCGAGCGGAGCGGUACAGUUUUAGAGCUAGAAAUAGCAAGUAAAAUA  
AGGCUAGUCCG (SEQ ID NO:9)

FIG. 1B is a schematic illustration showing a longer version of the sgRNA used to recruit Cas9 nuclease to a specific DNA sequence and to thereby introduce targeted alterations. The sequence of the guide RNA shown is  
GGAGCGAGCGGAGCGGUACAGUUUUAGAGCUAGAAAUAGCAAGUAAAAUA  
AGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU  
(SEQ ID NO:10).

FIG. 1C is a schematic illustration showing a Cas9 protein containing D10A and H840A mutations to render the nuclease portion of the protein catalytically inactive, fused to a transcriptional activation domain and recruited to a specific DNA sequence by a sgRNA. The sequence of the guide RNA shown is  
GGAGCGAGCGGAGCGGUACAGUUUUAGAGCUAGAAAUAGCAAGUAAAAUA  
AGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU  
(SEQ ID NO:10).

FIG. 1D is a schematic depicting recruitment of dCas9-VP64 fusion protein to a specific genomic target sequence by a chimeric sgRNA.

FIG. 1E is a diagram illustrating the positions and orientations of 16 sgRNAs targeted to the endogenous human VEGFA gene promoter. Small horizontal arrows represent the first 20 nts of the gRNA complementary to the genomic DNA sequence with the arrow pointing 5' to 3'. Grey bars indicate DNaseI hypersensitive sites previously defined in human 293 cells (Liu et al., J Biol Chem. 2001 Apr 6;276(14):11323-34), numbered relative to the transcription start site (right-angle arrow).

FIG. 2A is a bar graph showing activation of VEGFA protein expression in 293 cells by various sgRNAs, each expressed with (grey bars) or without (black bars) dCas9-VP64. Fold-activation of VEGFA was calculated relative to the off-target sgRNA control as described in Methods. Each experiment was performed in triplicate and error bars

represent standard errors of the mean. Asterisks indicate samples that are significantly elevated above the off-target control as determined by a paired, one-sided t-test ( $p < 0.05$ ).

FIG. 2B is a bar graph showing multiplex sgRNA expression induces synergistic activation of VEGFA protein expression by dCas9-VP64 protein. Fold-activation of VEGFA protein in 293 cells in which the indicated combinations of sgRNAs were co-expressed with dCas9-VP64 is shown. Note that in all of these experiments the amount of each individual sgRNA expression plasmid used for transfection was the same. Fold-activation values were calculated as described in 2A and shown as grey bars. The calculated sum of mean fold-activation values induced by individual sgRNAs is shown for each combination as black bars. Asterisks indicate all combinations that were found to be significantly greater than the expected sum as determined by an analysis of variance (ANOVA) ( $p < 0.05$ ).

FIG. 3A is a diagram illustrating the positions and orientations of six sgRNAs targeted to the endogenous human *NTF3* gene promoter. Horizontal arrows represent the first 20 nts of the sgRNA complementary to the genomic DNA sequence with the arrow pointing 5' to 3'. Grey line indicates region of potential open chromatin identified from the ENCODE DNaseI hypersensitivity track on the UCSC genome browser with the thicker part of the bar indicating the first transcribed exon. Numbering shown is relative to the transcription start site (+1, right-angle arrow).

FIG. 3B is a bar graph showing activation of *NTF3* gene expression by sgRNA-guided dCas9-VP64 in 293 cells. Relative expression of *NTF3* mRNA, detected by quantitative RT-PCR and normalized to a *GAPDH* control ( $\Delta Ct \times 10^4$ ), is shown for 293 cells co-transfected with the indicated amounts of dCas9-VP64 and *NTF3*-targeted sgRNA expression plasmids. All experiments were performed in triplicate with error bars representing standard errors of the mean. Asterisks indicate samples that are significantly greater than the off-target gRNA control as determined by a paired, one-sided T-test ( $P < 0.05$ ).

FIG. 3C is a bar graph showing multiplex gRNA expression induces synergistic activation of *NTF3* mRNA expression by dCas9-VP64 protein. Relative expression of *NTF3* mRNA, detected by quantitative RT-PCR and normalized to a *GAPDH* control ( $\Delta Ct \times 10^4$ ), is shown for 293 cells co-transfected with dCas9-VP64 and the indicated combinations of *NTF3*-targeted gRNA expression plasmids. Note that in all of these experiments the amount of each individual gRNA expression plasmid used for transfection was the same. All experiments were performed in triplicate with error bars representing

FIG. 4 is an exemplary sequence of an sgRNA expression vector.

FIG. 6 is an exemplary sequence of CMV-T7-Cas9 recoded D10A/H840A-3XFLAG-VP64 expression vector.

DALDDFDLMLGSDALDDFDLMLGSDALDDFDLMLGSDALDDFDLML (SEQ ID NO:12), is boxed.

FIG. 9 is an exemplary sequence of dCas9-TET1.

FIGS. 11A-B are bar graphs showing the activities of dCas9-VP64 activators bearing alternative substitution mutations to catalytically inactivate Cas9 function. (11A) Plasmids expressing dCas9-VP64 proteins bearing various Cas9 inactivating substitutions to residues D10 and H840 were each co-transfected into HEK293 cells with either a single gRNA or three distinctly targeted gRNAs targeting the VEGFA upstream region (blue and red bars, respectively). (11B) Plasmids expressing these dCas9-VP64 variants were also transfected into a HEK293 cell-line that stably expresses a single VEGFA-targeted gRNA.

VEGFA protein levels were determined by ELISA with mean of two replicates and standard errors of the mean (error bars) shown.

### DETAILED DESCRIPTION

Described herein are fusion proteins of a heterologous functional domain (e.g., a transcriptional activation domain) fused to a catalytically inactivated version of the Cas9 protein for the purpose of enabling RNA-guided targeting of these functional domains to specific genomic locations in cells and living organisms.

The CRISPR/Cas system has evolved in bacteria as a defense mechanism to protect against invading plasmids and viruses. Short protospacers, derived from foreign nucleic acid, are incorporated into CRISPR loci and subsequently transcribed and processed into short CRISPR RNAs (crRNAs). These crRNAs, complexed with a second tracrRNA, then use their sequence complementarity to the invading nucleic acid to guide Cas9-mediated cleavage, and consequent destruction of the foreign nucleic acid. In 2012, Doudna and colleagues demonstrated that a single guide RNA (sgRNA) composed of a fusion of a crRNA with tracrRNA can mediate recruitment of Cas9 nuclease to specific DNA sequences in vitro (Fig. 1C; Jinek et al., Science 2012).

More recently, a longer version of the sgRNA has been used to introduce targeted alterations in human cells and zebrafish (Fig. 1B; Mali et al. Science 2013, Hwang and Fu et al., Nat Biotechnol. 2013 Mar;31(3):227-9). Qi et al. demonstrated that gRNA-mediated recruitment of a catalytically inactive mutant form of Cas9 (referred to as dCas9) could lead to repression of specific endogenous genes in *E. coli* as well as of an EGFP reporter gene in human cells (Qi et al., Cell 152, 1173–1183 (2013)). Although this study demonstrated the potential to adapt RNA-guided Cas9 technology for regulation of gene expression, it did not test or demonstrate whether heterologous functional domains (e.g.—transcriptional activation domains) could be fused to dCas9 without disrupting its ability to be recruited to specific genomic sites by programmable sgRNAs or dual gRNAs (dgRNAs – i.e.- a customized crRNA and a tracrRNA).

As described herein, in addition to guiding Cas9-mediated nuclease activity, it is possible to use CRISPR-derived RNAs to target heterologous functional domains fused to Cas9 (Cas9-HFD) to specific sites in the genome (Figure 1C). For example, as described herein, it is possible to use single guide RNAs (sgRNAs) to target Cas9-HFD, e.g., Cas9-transcriptional activators (hereafter referred to as Cas9-activators) to the promoters of specific genes and thereby increase expression of the target gene. Thus Cas9-HFD can be

localized to sites in the genome, with target specificity defined by sequence complementarity of the guide RNA. The target sequence also includes a PAM sequence (a 2-5 nucleotide sequence specified by the Cas9 protein which is adjacent to the sequence specified by the RNA).

- 5 The Cas9-HFD are created by fusing a heterologous functional domain (e.g., a transcriptional activation domain, e.g., from VP64 or NF- $\kappa$ B p65), to the N-terminus or C-terminus of a catalytically inactive Cas9 protein.

### Cas9

- 10 A number of bacteria express Cas9 protein variants. The Cas9 from *Streptococcus pyogenes* is presently the most commonly used; some of the other Cas9 proteins have high levels of sequence identity with the *S. pyogenes* Cas9 and use the same guide RNAs. Others are more diverse, use different gRNAs, and recognize different PAM sequences as well (the 2-5 nucleotide sequence specified by the protein which is adjacent to the sequence specified by the RNA). Chylinski et al. classified Cas9 proteins from a large group of  
15 bacteria (RNA Biology 10:5, 1–12; 2013), and a large number of Cas9 proteins are listed in supplementary figure 1 and supplementary table 1 thereof, which are incorporated by reference herein. Additional Cas9 proteins are described in Esvelt et al., Nat Methods. 2013 Nov; 10(11):1116-21 and Fonfara et al., “Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems.”  
20 Nucleic Acids Res. 2013 Nov 22. [Epub ahead of print] doi:10.1093/nar/gkt1074.

- Cas9 molecules of a variety of species can be used in the methods and compositions described herein. While the *S. pyogenes* and *S. thermophilus* Cas9 molecules are the subject of much of the disclosure herein, Cas9 molecules of, derived from, or based on the Cas9 proteins of other species listed herein can be used as well. In other words, while the  
25 much of the description herein uses *S. pyogenes* and *S. thermophilus* Cas9 molecules, Cas9 molecules from the other species can replace them. Such species include those set forth in the following table, which was created based on supplementary figure 1 of Chylinski et al., 2013.

| Alternative Cas9 proteins |                                                        |
|---------------------------|--------------------------------------------------------|
| GenBank Acc No.           | Bacterium                                              |
| 303229466                 | <i>Veillonella atypica</i> ACS-134-V-Col7a             |
| 34762592                  | <i>Fusobacterium nucleatum</i> subsp. <i>vincentii</i> |
| 374307738                 | <i>Filifactor alocis</i> ATCC 35896                    |
| 320528778                 | <i>Solobacterium moorei</i> F0204                      |
| 291520705                 | <i>Coprococcus catus</i> GD-7                          |

| Alternative Cas9 proteins |                                                            |
|---------------------------|------------------------------------------------------------|
| GenBank Acc No.           | Bacterium                                                  |
| 42525843                  | <i>Treponema denticola</i> ATCC 35405                      |
| 304438954                 | <i>Peptoniphilus duerdenii</i> ATCC BAA-1640               |
| 224543312                 | <i>Catenibacterium mitsuokai</i> DSM 15897                 |
| 24379809                  | <i>Streptococcus mutans</i> UA159                          |
| 15675041                  | <i>Streptococcus pyogenes</i> SF370                        |
| 16801805                  | <i>Listeria innocua</i> Clip11262                          |
| 116628213                 | <i>Streptococcus thermophilus</i> LMD-9                    |
| 323463801                 | <i>Staphylococcus pseudintermedius</i> ED99                |
| 352684361                 | <i>Acidaminococcus intestini</i> RyC-MR95                  |
| 302336020                 | <i>Olsenella uli</i> DSM 7084                              |
| 366983953                 | <i>Oenococcus kitaharae</i> DSM 17330                      |
| 310286728                 | <i>Bifidobacterium bifidum</i> S17                         |
| 258509199                 | <i>Lactobacillus rhamnosus</i> GG                          |
| 300361537                 | <i>Lactobacillus gasseri</i> JV-V03                        |
| 169823755                 | <i>Finegoldia magna</i> ATCC 29328                         |
| 47458868                  | <i>Mycoplasma mobile</i> 163K                              |
| 284931710                 | <i>Mycoplasma gallisepticum</i> str. F                     |
| 363542550                 | <i>Mycoplasma ovipneumoniae</i> SC01                       |
| 384393286                 | <i>Mycoplasma canis</i> PG 14                              |
| 71894592                  | <i>Mycoplasma synoviae</i> 53                              |
| 238924075                 | <i>Eubacterium rectale</i> ATCC 33656                      |
| 116627542                 | <i>Streptococcus thermophilus</i> LMD-9                    |
| 315149830                 | <i>Enterococcus faecalis</i> TX0012                        |
| 315659848                 | <i>Staphylococcus lugdunensis</i> M23590                   |
| 160915782                 | <i>Eubacterium dolichum</i> DSM 3991                       |
| 336393381                 | <i>Lactobacillus coryniformis</i> subsp. <i>torquens</i>   |
| 310780384                 | <i>Ilyobacter polytropus</i> DSM 2926                      |
| 325677756                 | <i>Ruminococcus albus</i> 8                                |
| 187736489                 | <i>Akkermansia muciniphila</i> ATCC BAA-835                |
| 117929158                 | <i>Acidothermus cellulolyticus</i> 11B                     |
| 189440764                 | <i>Bifidobacterium longum</i> DJO10A                       |
| 283456135                 | <i>Bifidobacterium dentium</i> Bd1                         |
| 38232678                  | <i>Corynebacterium diphtheriae</i> NCTC 13129              |
| 187250660                 | <i>Elusimicrobium minutum</i> Pei191                       |
| 319957206                 | <i>Nitratifractor salsuginis</i> DSM 16511                 |
| 325972003                 | <i>Sphaerochaeta globus</i> str. <i>Buddy</i>              |
| 261414553                 | <i>Fibrobacter succinogenes</i> subsp. <i>succinogenes</i> |
| 60683389                  | <i>Bacteroides fragilis</i> NCTC 9343                      |
| 256819408                 | <i>Capnocytophaga ochracea</i> DSM 7271                    |
| 90425961                  | <i>Rhodopseudomonas palustris</i> BisB18                   |
| 373501184                 | <i>Prevotella micans</i> F0438                             |
| 294674019                 | <i>Prevotella ruminicola</i> 23                            |
| 365959402                 | <i>Flavobacterium columnare</i> ATCC 49512                 |
| 312879015                 | <i>Aminomonas paucivorans</i> DSM 12260                    |
| 83591793                  | <i>Rhodospirillum rubrum</i> ATCC 11170                    |
| 294086111                 | <i>Candidatus Puniceispirillum marinum</i> IMCC1322        |

| Alternative Cas9 proteins |                                                      |
|---------------------------|------------------------------------------------------|
| GenBank Acc No.           | Bacterium                                            |
| 121608211                 | <i>Verminephrobacter eiseniae</i> EF01-2             |
| 344171927                 | <i>Ralstonia syzygii</i> R24                         |
| 159042956                 | <i>Dinoroseobacter shibae</i> DFL 12                 |
| 288957741                 | <i>Azospirillum</i> sp- B510                         |
| 92109262                  | <i>Nitrobacter hamburgensis</i> X14                  |
| 148255343                 | <i>Bradyrhizobium</i> sp- BTai1                      |
| 34557790                  | <i>Wolinella succinogenes</i> DSM 1740               |
| 218563121                 | <i>Campylobacter jejuni</i> subsp. <i>jejuni</i>     |
| 291276265                 | <i>Helicobacter mustelae</i> 12198                   |
| 229113166                 | <i>Bacillus cereus</i> Rock1-15                      |
| 222109285                 | <i>Acidovorax ebreus</i> TPSY                        |
| 189485225                 | uncultured Termite group 1                           |
| 182624245                 | <i>Clostridium perfringens</i> D str.                |
| 220930482                 | <i>Clostridium cellulolyticum</i> H10                |
| 154250555                 | <i>Parvibaculum lavamentivorans</i> DS-1             |
| 257413184                 | <i>Roseburia intestinalis</i> L1-82                  |
| 218767588                 | <i>Neisseria meningitidis</i> Z2491                  |
| 15602992                  | <i>Pasteurella multocida</i> subsp. <i>multocida</i> |
| 319941583                 | <i>Sutterella wadsworthensis</i> 3 1                 |
| 254447899                 | <i>gamma proteobacterium</i> HTCC5015                |
| 54296138                  | <i>Legionella pneumophila</i> str. Paris             |
| 331001027                 | <i>Parasutterella excrementihominis</i> YIT 11859    |
| 34557932                  | <i>Wolinella succinogenes</i> DSM 1740               |
| 118497352                 | <i>Francisella novicida</i> U112                     |

The constructs and methods described herein can include the use of any of those Cas9 proteins, and their corresponding guide RNAs or other guide RNAs that are compatible. The Cas9 from *Streptococcus thermophilus* LMD-9 CRISPR1 system has been shown to function in human cells in Cong et al (Science 339, 819 (2013)). Additionally, Jinek et al. showed *in vitro* that Cas9 orthologs from *S. thermophilus* and *L. innocua*, (but not from *N. meningitidis* or *C. jejuni*, which likely use a different guide RNA), can be guided by a dual *S. pyogenes* gRNA to cleave target plasmid DNA, albeit with slightly decreased efficiency.

In some embodiments, the present system utilizes the Cas9 protein from *S. pyogenes*, either as encoded in bacteria or codon-optimized for expression in mammalian cells, containing mutations at D10, E762, H983, or D986 and H840 or N863, e.g., D10A/D10N and H840A/H840N/H840Y, to render the nuclease portion of the protein catalytically inactive; substitutions at these positions could be alanine (as they are in Nishimasu et al., Cell 156, 935–949 (2014)) or they could be other residues, e.g., glutamine, asparagine, tyrosine, serine, or aspartate, e.g., E762Q, H983N, H983Y, D986N, N863D, N863S, or N863H (Figure 1C). The sequence of the catalytically inactive *S. pyogenes* Cas9

that can be used in the methods and compositions described herein is as follows; the exemplary mutations of D10A and H840A are in bold and underlined.

|    |                           |             |            |            |            |                           |
|----|---------------------------|-------------|------------|------------|------------|---------------------------|
| 5  | 10                        | 20          | 30         | 40         | 50         | 60                        |
|    | MDKKYSIGL <b><u>A</u></b> | IGTNSVGWAV  | ITDEYKVPSK | KFKVLGNTDR | HSIKKNLIGA | LLFDSGETAE                |
| 10 | 70                        | 80          | 90         | 100        | 110        | 120                       |
|    | ATRLKRTARR                | RYTRRKNRIC  | YLQEIFSNE  | AKVDDSFHR  | LEESFLVEED | KKHERHPIFG                |
| 15 | 130                       | 140         | 150        | 160        | 170        | 180                       |
|    | NIVDEVAYHE                | KYPTIYHLRK  | KLVDSTDKAD | LRLIYLALAH | MIKFRGHFLI | EGDLNPDNSD                |
| 20 | 190                       | 200         | 210        | 220        | 230        | 240                       |
|    | VDKLFIQLVQ                | TYNQLFEENP  | INASGVDAKA | ILSARLSKSR | RLENLIAQLP | GEKKNGLFGN                |
| 25 | 250                       | 260         | 270        | 280        | 290        | 300                       |
|    | LIALSLGLTP                | NFKSNFDLAE  | DAKLQLSKDT | YDDDLNLLA  | QIGDQYADLF | LAAKNLSDAI                |
| 30 | 310                       | 320         | 330        | 340        | 350        | 360                       |
|    | LLSDILRVNT                | EITKAPLSAS  | MIKRYDEHHQ | DLTLLKALVR | QQLPEKYKEI | FFDQSKNGYA                |
| 35 | 370                       | 380         | 390        | 400        | 410        | 420                       |
|    | GYIDGGASQE                | EFYKFIKPIL  | EKMDGTEELL | VKLNREDLLR | KQRTFDNGSI | PHQIHLGELH                |
| 40 | 430                       | 440         | 450        | 460        | 470        | 480                       |
|    | AILRRQEDFY                | PFLKDNREKI  | EKILTFRIPY | YVGPLARGNS | RFAWMTRKSE | ETITPWNFEE                |
| 45 | 490                       | 500         | 510        | 520        | 530        | 540                       |
|    | VVDKGASAQ                 | FIERMNTNFDK | NLPNEKVLPK | HSLLYEYFTV | YNELTKVKYV | TEGMRKPAFL                |
| 50 | 550                       | 560         | 570        | 580        | 590        | 600                       |
|    | SGEQKKAIVD                | LLFKTNRKVT  | VKQLKEDYFK | KIECFDSVEI | SGVEDRFNAS | LGTYHDLKI                 |
| 55 | 610                       | 620         | 630        | 640        | 650        | 660                       |
|    | IKDKDFLDNE                | ENEDILEDIV  | LTLTLFEDRE | MIEERLKTYA | HLFDDKVMKQ | LKRRRYTGWG                |
| 60 | 670                       | 680         | 690        | 700        | 710        | 720                       |
|    | RLSRKLINGI                | RDQSGKTIL   | DFLKSDGFAN | RNFMQLIHDD | SLTFKEDIQK | AQVSGQGD                  |
| 65 | 730                       | 740         | 750        | 760        | 770        | 780                       |
|    | HEHIANLAGS                | PAIKKGILQT  | VKVVDLVKV  | MGRHKPENIV | IEMARENQTT | QKGQKNSRER                |
| 70 | 790                       | 800         | 810        | 820        | 830        | 840                       |
|    | MKRIEEGIKE                | LGSQILKEHP  | VENTQLQNEK | LYLYYLQNGR | DMYVDQELDI | NRLSDYD <b><u>VDA</u></b> |
| 75 | 850                       | 860         | 870        | 880        | 890        | 900                       |
|    | IVPQSFLKDD                | SIDNKVLTRS  | DKNRGKSDNV | PSEEVVKKMK | NYWRQLLNAK | LITQRKFDNL                |
| 80 | 910                       | 920         | 930        | 940        | 950        | 960                       |
|    | TKAERGGLSE                | LDKAGFIKRQ  | LVETRQITKH | VAQILDSRMN | TKYDENDKLI | REVKVITLKS                |
| 85 | 970                       | 980         | 990        | 1000       | 1010       | 1020                      |
|    | KLVSDFRKDF                | QFYKVRINN   | YHHAHDAYLN | AVVGTAIIK  | YPKLESEFVY | GDYKVYDVRK                |
| 90 | 1030                      | 1040        | 1050       | 1060       | 1070       | 1080                      |
|    | MIKSEQEIG                 | KATAKYFFYS  | NIMNFFKTEI | TLANGEIRKR | PLIETNGETG | EIVWDKGRDF                |
| 95 | 1090                      | 1100        | 1110       | 1120       | 1130       | 1140                      |
|    | ATVRKVLSP                 | QVNIVKKTEV  | QTGGFSKESI | LPKRNSDKLI | ARKKDWDPKK | YGGFDSPTVA                |

```

      1150      1160      1170      1180      1190      1200
YSVLVVAKVE KGKSKKLKSV KELLGITIME RSSFEKNPID FLEAKGYKEV KKDLI IKLPK

      1210      1220      1230      1240      1250      1260
5  YSLFELENGR KRMLASAGEL QKGNELALPS KYVNFLYLAS HYEKLKGSPE DNEQKQLFVE

      1270      1280      1290      1300      1310      1320
QHKHYLDEII EQISEFSKRV ILADANLDKV LSAYNKHHRDK PIREQAENII HLFTLTNLGA

10      1330      1340      1350      1360
PAAFKYFDTT IDRKYRTSTK EVLDATLIHQ SITGLYETRI DLSQLGGD (SEQ ID NO:13)

```

In some embodiments, the Cas9 nuclease used herein is at least about 50% identical to the sequence of *S. pyogenes* Cas9, i.e., at least 50% identical to SEQ ID NO:13. In some  
 15 embodiments, the nucleotide sequences are about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% identical to SEQ ID NO:13.

In some embodiments, the catalytically inactive Cas9 used herein is at least about 50% identical to the sequence of the catalytically inactive *S. pyogenes* Cas9, i.e., at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% identical to SEQ  
 20 ID NO:13, wherein the mutations at D10 and H840, e.g., D10A/D10N and H840A/H840N/H840Y are maintained.

In some embodiments, any differences from SEQ ID NO:13 are in non-conserved regions, as identified by sequence alignment of sequences set forth in Chylinski et al., RNA Biology 10:5, 1–12; 2013 (e.g., in supplementary figure 1 and supplementary table 1  
 25 thereof); Esvelt et al., Nat Methods. 2013 Nov;10(11):1116-21 and Fonfara et al., Nucl. Acids Res. (2014) 42 (4): 2577-2590. [Epub ahead of print 2013 Nov 22] doi:10.1093/nar/gkt1074, and wherein the mutations at D10 and H840, e.g., D10A/D10N and H840A/H840N/H840Y are maintained.

To determine the percent identity of two sequences, the sequences are aligned for  
 30 optimal comparison purposes (gaps are introduced in one or both of a first and a second amino acid or nucleic acid sequence as required for optimal alignment, and non-homologous sequences can be disregarded for comparison purposes). The length of a reference sequence aligned for comparison purposes is at least 50% (in some embodiments, about 50%, 55%, 60%, 65%, 70%, 75%, 85%, 90%, 95%, or 100% of the length of the  
 35 reference sequence) is aligned. The nucleotides or residues at corresponding positions are then compared. When a position in the first sequence is occupied by the same nucleotide or residue as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number

of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. For purposes of the present application, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch ((1970) J. Mol. Biol. 48:444-453) algorithm which has been incorporated into the GAP program in the GCG software package, using a Blossum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

## 10 Heterologous Functional Domains

The transcriptional activation domains can be fused on the N or C terminus of the Cas9. In addition, although the present description exemplifies transcriptional activation domains, other heterologous functional domains (e.g., transcriptional repressors (e.g., KRAB, ERD, SID, and others, e.g., amino acids 473–530 of the *ets2* repressor factor (ERF) repressor domain (ERD), amino acids 1–97 of the KRAB domain of KOX1, or amino acids 1–36 of the Mad mSIN3 interaction domain (SID); see Beerli et al., PNAS USA 95:14628-14633 (1998)) or silencers such as Heterochromatin Protein 1 (HP1, also known as swi6), e.g., HP1 $\alpha$  or HP1 $\beta$ ; proteins or peptides that could recruit long non-coding RNAs (lncRNAs) fused to a fixed RNA binding sequence such as those bound by the MS2 coat protein, endoribonuclease Csy4, or the lambda N protein; enzymes that modify the methylation state of DNA (e.g., DNA methyltransferase (DNMT) or TET proteins); or enzymes that modify histone subunits (e.g., histone acetyltransferases (HAT), histone deacetylases (HDAC), histone methyltransferases (e.g., for methylation of lysine or arginine residues) or histone demethylases (e.g., for demethylation of lysine or arginine residues)) as are known in the art can also be used. A number of sequences for such domains are known in the art, e.g., a domain that catalyzes hydroxylation of methylated cytosines in DNA. Exemplary proteins include the Ten-Eleven-Translocation (TET)1-3 family, enzymes that converts 5-methylcytosine (5-mC) to 5-hydroxymethylcytosine (5-hmC) in DNA.

Sequences for human TET1-3 are known in the art and are shown in the following table:

| Gene  | GenBank Accession Nos. |                |
|-------|------------------------|----------------|
|       | Amino Acid             | Nucleic Acid   |
| TET1  | NP_085128.2            | NM_030625.2    |
| TET2* | NP_001120680.1 (var 1) | NM_001127208.2 |
|       | NP_060098.3 (var 2)    | NM_017628.4    |
| TET3  | NP_659430.1            | NM_144993.1    |

\* Variant (1) represents the longer transcript and encodes the longer isoform (a). Variant (2) differs in the 5' UTR and in the 3' UTR and coding sequence compared to variant 1. The resulting isoform (b) is shorter and has a distinct C-terminus compared to isoform a.

In some embodiments, all or part of the full-length sequence of the catalytic domain can be included, e.g., a catalytic module comprising the cysteine-rich extension and the 2OGFeDO domain encoded by 7 highly conserved exons, e.g., the Tet1 catalytic domain comprising amino acids 1580-2052, Tet2 comprising amino acids 1290-1905 and Tet3 comprising amino acids 966-1678. See, e.g., Fig. 1 of Iyer et al., Cell Cycle. 2009 Jun 1;8(11):1698-710. Epub 2009 Jun 27, for an alignment illustrating the key catalytic residues in all three Tet proteins, and the supplementary materials thereof (available at ftp site [ftp.ncbi.nih.gov/pub/aravind/DONS/supplementary\\_material\\_DONS.html](ftp.ncbi.nih.gov/pub/aravind/DONS/supplementary_material_DONS.html)) for full length sequences (see, e.g., seq 2c); in some embodiments, the sequence includes amino acids 1418-2136 of Tet1 or the corresponding region in Tet2/3.

Other catalytic modules can be from the proteins identified in Iyer et al., 2009.

In some embodiments, the heterologous functional domain is a biological tether, and comprises all or part of (e.g., DNA binding domain from) the MS2 coat protein, endoribonuclease Csy4, or the lambda N protein. These proteins can be used to recruit RNA molecules containing a specific stem-loop structure to a locale specified by the dCas9 gRNA targeting sequences. For example, a dCas9 fused to MS2 coat protein, endoribonuclease Csy4, or lambda N can be used to recruit a long non-coding RNA (lncRNA) such as XIST or HOTAIR; see, e.g., Keryer-Bibens et al., Biol. Cell 100:125-138 (2008), that is linked to the Csy4, MS2 or lambda N binding sequence. Alternatively, the Csy4, MS2 or lambda N protein binding sequence can be linked to another protein, e.g., as described in Keryer-Bibens et al., supra, and the protein can be targeted to the dCas9

binding site using the methods and compositions described herein. In some embodiments, the Csy4 is catalytically inactive.

In some embodiments, the fusion proteins include a linker between the dCas9 and the heterologous functional domains. Linkers that can be used in these fusion proteins (or  
5 between fusion proteins in a concatenated structure) can include any sequence that does not interfere with the function of the fusion proteins. In preferred embodiments, the linkers are short, e.g., 2-20 amino acids, and are typically flexible (i.e., comprising amino acids with a high degree of freedom such as glycine, alanine, and serine). In some embodiments, the linker comprises one or more units consisting of GGGS (SEQ ID NO:14) or GGGGS (SEQ  
10 ID NO:15), e.g., two, three, four, or more repeats of the GGGS (SEQ ID NO:14) or GGGGS (SEQ ID NO:15) unit. Other linker sequences can also be used.

### Methods of Use

The described Cas9-HFD system is a useful and versatile tool for modifying the expression of endogenous genes. Current methods for achieving this require the generation  
15 of novel engineered DNA-binding proteins (such as engineered zinc finger or transcription activator-like effector DNA binding domains) for each site to be targeted. Because these methods demand expression of a large protein specifically engineered to bind each target site, they are limited in their capacity for multiplexing. Cas9-HFD, however, require expression of only a single Cas9-HFD protein, which can be targeted to multiple sites in the  
20 genome by expression of multiple short gRNAs. This system could therefore easily be used to simultaneously induce expression of a large number of genes or to recruit multiple Cas9-HFDs to a single gene, promoter, or enhancer. This capability will have broad utility, e.g., for basic biological research, where it can be used to study gene function and to manipulate the expression of multiple genes in a single pathway, and in synthetic biology, where it will  
25 enable researchers to create circuits in cell that are responsive to multiple input signals. The relative ease with which this technology can be implemented and adapted to multiplexing will make it a broadly useful technology with many wide-ranging applications.

The methods described herein include contacting cells with a nucleic acid encoding  
30 the Cas9-HFD described herein, and nucleic acids encoding one or more guide RNAs directed to a selected gene, to thereby modulate expression of that gene.

*Guide RNAs (gRNAs)*

Guide RNAs generally speaking come in two different systems: System 1, which uses separate crRNA and tracrRNAs that function together to guide cleavage by Cas9, and System 2, which uses a chimeric crRNA-tracrRNA hybrid that combines the two separate guide RNAs in a single system (referred to as a single guide RNA or sgRNA, see also Jinek et al., Science 2012; 337:816–821). The tracrRNA can be variably truncated and a range of lengths has been shown to function in both the separate system (system 1) and the chimeric gRNA system (system 2). For example, in some embodiments, tracrRNA may be truncated from its 3' end by at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35 or 40 nts. In some embodiments, the tracrRNA molecule may be truncated from its 5' end by at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35 or 40 nts. Alternatively, the tracrRNA molecule may be truncated from both the 5' and 3' end, e.g., by at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 or 20 nts on the 5' end and at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35 or 40 nts on the 3' end. See, e.g., Jinek et al., Science 2012; 337:816–821; Mali et al., Science. 2013 Feb 15;339(6121):823-6; Cong et al., Science. 2013 Feb 15;339(6121):819-23; and Hwang and Fu et al., Nat Biotechnol. 2013 Mar;31(3):227-9; Jinek et al., Elife 2, e00471 (2013)). For System 2, generally the longer length chimeric gRNAs have shown greater on-target activity but the relative specificities of the various length gRNAs currently remain undefined and therefore it may be desirable in certain instances to use shorter gRNAs. In some embodiments, the gRNAs are complementary to a region that is within about 100-800 bp upstream of the transcription start site, e.g., is within about 500 bp upstream of the transcription start site, includes the transcription start site, or within about 100-800 bp, e.g., within about 500 bp, downstream of the transcription start site. In some embodiments, vectors (e.g., plasmids) encoding more than one gRNA are used, e.g., plasmids encoding, 2, 3, 4, 5, or more gRNAs directed to different sites in the same region of the target gene.

Cas9 nuclease can be guided to specific 17-20 nt genomic targets bearing an additional proximal protospacer adjacent motif (PAM), e.g., of sequence NGG, using a guide RNA, e.g., a single gRNA or a tracrRNA/crRNA, bearing 17-20 nts at its 5' end that are complementary to the complementary strand of the genomic DNA target site. Thus, the present methods can include the use of a single guide RNA comprising a crRNA fused to a normally trans-encoded tracrRNA, e.g., a single Cas9 guide RNA as described in Mali et al., Science 2013 Feb 15; 339(6121):823-6, with a sequence at the 5' end that is complementary to the target sequence, e.g., of 25-17, optionally 20 or fewer nucleotides (nts), e.g., 20, 19, 18, or 17 nts, preferably 17 or 18 nts, of the complementary strand to a

target sequence immediately 5' of a protospacer adjacent motif (PAM), e.g., NGG, NAG, or NNGG. In some embodiments, the single Cas9 guide RNA consists of the sequence:

(X<sub>17-20</sub>)GUUUUAGAGCUAGAAAUAGCAAGUUA AAAUAAGGCUAGUCCG(X<sub>N</sub>)  
(SEQ ID NO:1);

5 (X<sub>17-20</sub>)GUUUUAGAGCUAUGCUGAAAAGCAUAGCAAGUUA AAAUAAGGCUAGUCCGUUAUC(X<sub>N</sub>) (SEQ ID NO:2);

(X<sub>17-20</sub>)GUUUUAGAGCUAUGCUGUUUUGGAAACAAAACAGCAUAGCAAGUUA  
AAUAAGGCUAGUCCGUUAUC(X<sub>N</sub>) (SEQ ID NO:3);

(X<sub>17-20</sub>)GUUUUAGAGCUAGAAAUAGCAAGUUA AAAUAAGGCUAGUCCGUUAUC  
10 AACUUGAAAAAGUGGCACCGAGUCGGUGC(X<sub>N</sub>) (SEQ ID NO:4),

(X<sub>17-20</sub>)GUUUUAGAGCUAGAAAUAGCAAGUUA AAAUAAGGCUAGUCCGUUAUC  
AACUUGAAAAAGUGGCACCGAGUCGGUGC(SEQ ID NO:5);

(X<sub>17-20</sub>)GUUUUAGAGCUAUGCUGGAAACAGCAUAGCAAGUUA AAAUAAGGCUA  
GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO:6); or

15 (X<sub>17-20</sub>)GUUUUAGAGCUAUGCUGGAAACAGCAUAGCAAGUUA AAAUAAGGCUA  
GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO:7);

wherein X<sub>17-20</sub> is the nucleotide sequence complementary to 17-20 consecutive nucleotides of the target sequence. DNAs encoding the single guide RNAs have been described previously in the literature (Jinek et al., Science. 337(6096):816-21 (2012) and Jinek et al.,  
20 Elife. 2:e00471 (2013)).

The guide RNAs can include X<sub>N</sub> which can be any sequence, wherein N (in the RNA) can be 0-200, e.g., 0-100, 0-50, or 0-20, that does not interfere with the binding of the ribonucleic acid to Cas9.

In some embodiments, the guide RNA includes one or more Adenine (A) or Uracil  
25 (U) nucleotides on the 3' end. In some embodiments the RNA includes one or more U, e.g., 1 to 8 or more Us (e.g., U, UU, UUU, UUUU, UUUUU, UUUUUU, UUUUUUU, UUUUUUUU) at the 3' end of the molecule, as a result of the optional presence of one or more Ts used as a termination signal to terminate RNA PolIII transcription.

Although some of the examples described herein utilize a single gRNA, the methods  
30 can also be used with dual gRNAs (e.g., the crRNA and tracrRNA found in naturally occurring systems). In this case, a single tracrRNA would be used in conjunction with multiple different crRNAs expressed using the present system, e.g., the following:

(X<sub>17-20</sub>)GUUUUAGAGCUA (SEQ ID NO:102);

(X<sub>17-20</sub>) GUUUUAGAGCUAUGCUGUUUUG (SEQ ID NO:103); or

(X<sub>17-20</sub>)GUUUUAGAGCUAUGCU (SEQ ID NO:104); and a tracrRNA sequence. In this case, the crRNA is used as the guide RNA in the methods and molecules described herein, and the tracrRNA can be expressed from the same or a different DNA molecule. In some embodiments, the methods include contacting the cell with a tracrRNA comprising or

consisting of the sequence

GGAACCAUUCAAAACAGCAUAGCAAGUUA AAAUAAGGCUAGUCCGUUAUCA  
ACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO:8) or an active portion  
thereof (an active portion is one that retains the ability to form complexes with Cas9 or dCas9). In some embodiments, the tracrRNA molecule may be truncated from its 3' end by  
at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35 or 40 nts. In another embodiment, the  
tracrRNA molecule may be truncated from its 5' end by at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10,  
15, 20, 25, 30, 35 or 40 nts. Alternatively, the tracrRNA molecule may be truncated from  
both the 5' and 3' end, e.g., by at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 or 20 nts on the 5' end  
and at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35 or 40 nts on the 3' end.

Exemplary tracrRNA sequences in addition to SEQ ID NO:8 include the following:

UAGCAAGUUA AAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCG  
AGUCGGUGC (SEQ ID NO:105) or an active portion thereof; or  
AGCAUAGCAAGUUA AAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGC  
ACCGAGUCGGUGC (SEQ ID NO:106) or an active portion thereof.

In some embodiments when (X<sub>17-20</sub>)GUUUUAGAGCUAUGCUGUUUUG (SEQ ID NO:102) is used as a crRNA, the following tracrRNA is used:

GGAACCAUUCAAAACAGCAUAGCAAGUUA AAAUAAGGCUAGUCCGUUAUCA  
ACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO:8) or an active portion  
thereof.

In some embodiments when (X<sub>17-20</sub>)GUUUUAGAGCUA (SEQ ID NO:102) is used as a crRNA, the following tracrRNA is used:

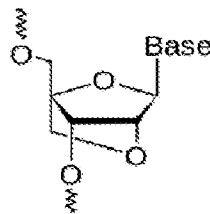
UAGCAAGUUA AAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCG  
AGUCGGUGC (SEQ ID NO:105) or an active portion thereof.

In some embodiments when (X<sub>17-20</sub>) GUUUUAGAGCUAUGCU (SEQ ID NO:104) is used as a crRNA, the following tracrRNA is used:

AGCAUAGCAAGUUA AAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGC  
ACCGAGUCGGUGC (SEQ ID NO:106) or an active portion thereof.

In some embodiments, the gRNA is targeted to a site that is at least three or more mismatches different from any sequence in the rest of the genome in order to minimize off-target effects.

Modified RNA oligonucleotides such as locked nucleic acids (LNAs) have been demonstrated to increase the specificity of RNA-DNA hybridization by locking the modified oligonucleotides in a more favorable (stable) conformation. For example, 2'-O-methyl RNA is a modified base where there is an additional covalent linkage between the 2' oxygen and 4' carbon which when incorporated into oligonucleotides can improve overall thermal stability and selectivity (**Formula I**).



Formula I – Locked Nucleic Acid

Thus in some embodiments, the tru-gRNAs disclosed herein may comprise one or more modified RNA oligonucleotides. For example, the truncated guide RNAs molecules described herein can have one, some or all of the region of the guideRNA complementary to the target sequence are modified, e.g., locked (2'-O-4'-C methylene bridge), 5'-methylcytidine, 2'-O-methyl-pseudouridine, or in which the ribose phosphate backbone has been replaced by a polyamide chain (peptide nucleic acid), e.g., a synthetic ribonucleic acid.

In other embodiments, one, some or all of the nucleotides of the tru-gRNA sequence may be modified, e.g., locked (2'-O-4'-C methylene bridge), 5'-methylcytidine, 2'-O-methyl-pseudouridine, or in which the ribose phosphate backbone has been replaced by a polyamide chain (peptide nucleic acid), e.g., a synthetic ribonucleic acid.

In some embodiments, the single guide RNAs and/or crRNAs and/or tracrRNAs can include one or more Adenine (A) or Uracil (U) nucleotides on the 3' end.

Existing Cas9-based RGNs use gRNA-DNA heteroduplex formation to guide targeting to genomic sites of interest. However, RNA-DNA heteroduplexes can form a more promiscuous range of structures than their DNA-DNA counterparts. In effect, DNA-DNA duplexes are more sensitive to mismatches, suggesting that a DNA-guided nuclease may not bind as readily to off-target sequences, making them comparatively more specific than RNA-guided nucleases. Thus, the guide RNAs usable in the methods described herein

can be hybrids, i.e., wherein one or more deoxyribonucleotides, e.g., a short DNA oligonucleotide, replaces all or part of the gRNA, e.g., all or part of the complementarity region of a gRNA. This DNA-based molecule could replace either all or part of the gRNA in a single gRNA system or alternatively might replace all or part of the crRNA and/or  
5 tracrRNA in a dual crRNA/tracrRNA system. Such a system that incorporates DNA into the complementarity region should more reliably target the intended genomic DNA sequences due to the general intolerance of DNA-DNA duplexes to mismatching compared to RNA-DNA duplexes. Methods for making such duplexes are known in the art, See, e.g., Barker et al., BMC Genomics. 2005 Apr 22;6:57; and Sugimoto et al., Biochemistry. 2000  
10 Sep 19;39(37):11270-81.

In addition, in a system that uses separate crRNA and tracrRNA, one or both can be synthetic and include one or more modified (e.g., locked) nucleotides or deoxyribonucleotides.

In a cellular context, complexes of Cas9 with these synthetic gRNAs could be used  
15 to improve the genome-wide specificity of the CRISPR/Cas9 nuclease system.

The methods described can include expressing in a cell, or contacting the cell with, a Cas9 gRNA plus a fusion protein as described herein.

### **Expression Systems**

In order to use the fusion proteins and guide RNAs described herein, it may be  
20 desirable to express them from a nucleic acid that encodes them. This can be performed in a variety of ways. For example, a nucleic acid encoding a guide RNA or fusion protein can be cloned into an intermediate vector for transformation into prokaryotic or eukaryotic cells for replication and/or expression. Intermediate vectors are typically prokaryote vectors, e.g., plasmids, or shuttle vectors, or insect vectors, for storage or manipulation of the  
25 nucleic acid encoding the fusion protein or for production of the fusion protein. The nucleic acid encoding the guide RNA or fusion protein can also be cloned into an expression vector, for administration to a plant cell, animal cell, preferably a mammalian cell or a human cell, fungal cell, bacterial cell, or protozoan cell.

To obtain expression, a sequence encoding a guide RNA or fusion protein is  
30 typically subcloned into an expression vector that contains a promoter to direct transcription. Suitable bacterial and eukaryotic promoters are well known in the art and described, e.g., in Sambrook et al., Molecular Cloning, A Laboratory Manual (3d ed. 2001); Kriegler, Gene Transfer and Expression: A Laboratory Manual (1990); and Current

Protocols in Molecular Biology (Ausubel et al., eds., 2010). Bacterial expression systems for expressing the engineered protein are available in, e.g., *E. coli*, *Bacillus* sp., and *Salmonella* (Palva et al., 1983, Gene 22:229-235). Kits for such expression systems are commercially available. Eukaryotic expression systems for mammalian cells, yeast, and insect cells are well known in the art and are also commercially available.

The promoter used to direct expression of the nucleic acid depends on the particular application. For example, a strong constitutive promoter is typically used for expression and purification of fusion proteins. In contrast, when the fusion protein is to be administered in vivo for gene regulation, either a constitutive or an inducible promoter can be used, depending on the particular use of the fusion protein. In addition, a preferred promoter for administration of the fusion protein can be a weak promoter, such as HSV TK or a promoter having similar activity. The promoter can also include elements that are responsive to transactivation, e.g., hypoxia response elements, Gal4 response elements, lac repressor response element, and small molecule control systems such as tetracycline-regulated systems and the RU-486 system (see, e.g., Gossen & Bujard, 1992, Proc. Natl. Acad. Sci. USA, 89:5547; Oligino et al., 1998, Gene Ther., 5:491-496; Wang et al., 1997, Gene Ther., 4:432-441; Neering et al., 1996, Blood, 88:1147-55; and Rendahl et al., 1998, Nat. Biotechnol., 16:757-761).

In addition to the promoter, the expression vector typically contains a transcription unit or expression cassette that contains all the additional elements required for the expression of the nucleic acid in host cells, either prokaryotic or eukaryotic. A typical expression cassette thus contains a promoter operably linked, e.g., to the nucleic acid sequence encoding the fusion protein, and any signals required, e.g., for efficient polyadenylation of the transcript, transcriptional termination, ribosome binding sites, or translation termination. Additional elements of the cassette may include, e.g., enhancers, and heterologous spliced intronic signals.

The particular expression vector used to transport the genetic information into the cell is selected with regard to the intended use of the fusion protein, e.g., expression in plants, animals, bacteria, fungus, protozoa, etc. Standard bacterial expression vectors include plasmids such as pBR322 based plasmids, pSKF, pET23D, and commercially available tag-fusion expression systems such as GST and LacZ. A preferred tag-fusion protein is the maltose binding protein (MBP). Such tag-fusion proteins can be used for purification of the engineered TALE repeat protein. Epitope tags can also be added to

recombinant proteins to provide convenient methods of isolation, for monitoring expression, and for monitoring cellular and subcellular localization, e.g., c-myc or FLAG.

Expression vectors containing regulatory elements from eukaryotic viruses are often used in eukaryotic expression vectors, e.g., SV40 vectors, papilloma virus vectors, and  
5 vectors derived from Epstein-Barr virus. Other exemplary eukaryotic vectors include pMSG, pAV009/A+, pMTO10/A+, pMAMneo-5, baculovirus pDSVE, and any other vector allowing expression of proteins under the direction of the SV40 early promoter, SV40 late promoter, metallothionein promoter, murine mammary tumor virus promoter, Rous sarcoma virus promoter, polyhedrin promoter, or other promoters shown effective for expression in  
10 eukaryotic cells.

The vectors for expressing the guide RNAs can include RNA Pol III promoters to drive expression of the guide RNAs, e.g., the H1, U6 or 7SK promoters. These human promoters allow for expression of gRNAs in mammalian cells following plasmid transfection. Alternatively, a T7 promoter may be used, e.g., for in vitro transcription, and  
15 the RNA can be transcribed in vitro and purified. Vectors suitable for the expression of short RNAs, e.g., siRNAs, shRNAs, or other small RNAs, can be used.

Some expression systems have markers for selection of stably transfected cell lines such as thymidine kinase, hygromycin B phosphotransferase, and dihydrofolate reductase. High yield expression systems are also suitable, such as using a baculovirus vector in insect cells,  
20 with the fusion protein encoding sequence under the direction of the polyhedrin promoter or other strong baculovirus promoters.

The elements that are typically included in expression vectors also include a replicon that functions in *E. coli*, a gene encoding antibiotic resistance to permit selection of bacteria that harbor recombinant plasmids, and unique restriction sites in nonessential  
25 regions of the plasmid to allow insertion of recombinant sequences.

Standard transfection methods are used to produce bacterial, mammalian, yeast or insect cell lines that express large quantities of protein, which are then purified using standard techniques (see, e.g., Colley et al., 1989, J. Biol. Chem., 264:17619-22; Guide to Protein Purification, in Methods in Enzymology, vol. 182 (Deutscher, ed., 1990)).

Transformation of eukaryotic and prokaryotic cells are performed according to standard techniques (see, e.g., Morrison, 1977, J. Bacteriol. 132:349-351; Clark-Curtiss & Curtiss, Methods in Enzymology 101:347-362 (Wu et al., eds, 1983)).  
30

Any of the known procedures for introducing foreign nucleotide sequences into host cells may be used. These include the use of calcium phosphate transfection, polybrene,

protoplast fusion, electroporation, nucleofection, liposomes, microinjection, naked DNA, plasmid vectors, viral vectors, both episomal and integrative, and any of the other well-known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (see, e.g., Sambrook et al., supra). It is only  
5 necessary that the particular genetic engineering procedure used be capable of successfully introducing at least one gene into the host cell capable of expressing the protein of choice.

In some embodiments, the fusion protein includes a nuclear localization domain which provides for the protein to be translocated to the nucleus. Several nuclear localization sequences (NLS) are known, and any suitable NLS can be used. For example,  
10 many NLSs have a plurality of basic amino acids, referred to as a bipartite basic repeats (reviewed in Garcia-Bustos et al, 1991, Biochim. Biophys. Acta, 1071:83-101). An NLS containing bipartite basic repeats can be placed in any portion of chimeric protein and results in the chimeric protein being localized inside the nucleus. In preferred embodiments a nuclear localization domain is incorporated into the final fusion protein, as the ultimate  
15 functions of the fusion proteins described herein will typically require the proteins to be localized in the nucleus. However, it may not be necessary to add a separate nuclear localization domain in cases where the DBD domain itself, or another functional domain within the final chimeric protein, has intrinsic nuclear translocation function.

The present invention includes the vectors and cells comprising the vectors.

## 20 EXAMPLES

The invention is further described in the following examples, which do not limit the scope of the invention described in the claims.

### **Example 1. Engineering CRISPR/Cas Activator System:**

It was hypothesized that RNA-guided transcriptional activators could be created by  
25 fusing the strong synthetic VP64 activation domain (Beerli et al., Proc Natl Acad Sci USA 95, 14628–14633 (1998)) to the carboxy-terminus of the catalytically inactivated dCas9 protein (Fig. 1D).

To express guide RNAs (gRNAs) in human cells, a vector was engineered that would express the full length chimeric gRNA (a fusion of crRNA and tracrRNA originally  
30 described by Jinek et al. (Science 2012)) driven by a U6 promoter. Construction of the gRNA expression plasmids was performed as follows. Pairs of DNA oligonucleotides

encoding the variable 20 nt gRNA targeting sequences were annealed together to generate short double-strand DNA fragments with 4bp overhangs (Table 1).

| <b>Table 1. VEGFA and NTF3 gene target sites and associated oligonucleotides used to construct gRNA expression plasmids.</b> |                                    |                   |
|------------------------------------------------------------------------------------------------------------------------------|------------------------------------|-------------------|
| <b>gRNA</b>                                                                                                                  | <b>Target Site (including PAM)</b> | <b>SEQ ID NO:</b> |
| V1                                                                                                                           | GTGTGCAGACGGCAGTCACTAGG            | 16.               |
| V2                                                                                                                           | GAGCAGCGTCTTCGAGAGTGAGG            | 17.               |
| V3                                                                                                                           | GGTGAGTGAGTGTGTGCGTGTGG            | 18.               |
| V4                                                                                                                           | GTTGGAGCGGGGAGAAGGCCAGG            | 19.               |
| V5                                                                                                                           | GGGTGGGGGAGTTTGCTCCTGG             | 20.               |
| V6                                                                                                                           | GGCTTTGGAAAGGGGTGGGGGG             | 21.               |
| V7                                                                                                                           | GGGGCGGGGTCCCGGCGGGGCGG            | 22.               |
| V8                                                                                                                           | GCTCGGAGGTCTGTGGCGCTGGGG           | 23.               |
| V9                                                                                                                           | GACTCACCGGCCAGGGCGCTCGG            | 24.               |
| V10                                                                                                                          | GGCGCAGCGGTTAGGTGGACCGG            | 25.               |
| V11                                                                                                                          | GGCGCATGGCTCCGCCCCGCGG             | 26.               |
| V12                                                                                                                          | GCCACGACCTCCGAGCTACCCGG            | 27.               |
| V13                                                                                                                          | GCGGCGTGAGCCCTCCCCCTTGG            | 28.               |
| V14                                                                                                                          | GGAGGCGGGGTGGAGGGGGTCGG            | 29.               |
| V15                                                                                                                          | GGGCTCACGCCGCGCTCCGGCGG            | 30.               |
| V16                                                                                                                          | GACCCCTCCACCCGCGCTCCGG             | 31.               |
| N1                                                                                                                           | GAGCGCGGAGCCATCTGGCCGGG            | 32.               |
| N2                                                                                                                           | GCGCGGCGCGGAAGGGGTTAAGG            | 33.               |
| N3                                                                                                                           | GCGGCGGCGCGGGGCCGCGGG              | 34.               |
| N4                                                                                                                           | GCCGCGCGCCCTCCCCGCGCGG             | 35.               |
| N5                                                                                                                           | GCGGTTATAACCAGCCAACCCGG            | 36.               |
| N6                                                                                                                           | GTGCGCGGAGCTGTTTCGGAAGG            | 37.               |
| <b>gRNA</b>                                                                                                                  | <b>top oligo</b>                   | <b>SEQ ID NO:</b> |
| V1                                                                                                                           | ACACCGTGTGCAGACGGCAGTCACTG         | 38.               |
| V2                                                                                                                           | ACACCGAGCAGCGTCTTCGAGAGTGG         | 39.               |
| V3                                                                                                                           | ACACCGGTGAGTGAGTGTGTGCGTGG         | 40.               |
| V4                                                                                                                           | ACACCGTTGGAGCGGGGAGAAGGCCG         | 41.               |
| V5                                                                                                                           | ACACCGGGTGGGGGAGTTTGCTCCG          | 42.               |
| V6                                                                                                                           | ACACCGGCTTTGGAAAGGGGTGGGG          | 43.               |
| V7                                                                                                                           | ACACCGGGGCGGGGTCCCGGCGGGGG         | 44.               |
| V8                                                                                                                           | ACACCGCTCGGAGGTCTGTGGCGCTGG        | 45.               |
| V9                                                                                                                           | ACACCGACTCACCGGCCAGGGCGCTG         | 46.               |
| V10                                                                                                                          | ACACCGGCGCAGCGGTTAGGTGGACG         | 47.               |
| V11                                                                                                                          | ACACCGGCGCATGGCTCCGCCCCGCG         | 48.               |
| V12                                                                                                                          | ACACCGCCACGACCTCCGAGCTACCG         | 49.               |
| V13                                                                                                                          | ACACCGCGGCGTGAGCCCTCCCCCTG         | 50.               |
| V14                                                                                                                          | ACACCGGAGGCGGGGTGGAGGGGGTG         | 51.               |
| V15                                                                                                                          | ACACCGGGCTCACGCCGCGCTCCGGG         | 52.               |
| V16                                                                                                                          | ACACCGACCCCTCCACCCGCGCTCG          | 53.               |
| N1                                                                                                                           | ACACCGAGCGCGGAGCCATCTGGCCG         | 54.               |
| N2                                                                                                                           | ACACCGCGCGGCGCGGAAGGGGTTAG         | 55.               |
| N3                                                                                                                           | ACACCGCGGCGCGGCGGGGCCGCGG          | 56.               |
| N4                                                                                                                           | ACACCGCGCGCGCCCTCCCCGCGG           | 57.               |
| N5                                                                                                                           | ACACCGCGGTTATAACCAGCCAACCG         | 58.               |
| N6                                                                                                                           | ACACCGTGCGCGGAGCTGTTTCGGAAG        | 59.               |
| <b>gRNA</b>                                                                                                                  | <b>bottom oligo</b>                | <b>SEQ ID NO:</b> |
| V1                                                                                                                           | AAAACAGTGACTGCCGTCTGCACACG         | 60.               |
| V2                                                                                                                           | AAAACCACCTCTCGAAGACGCTGCTCG        | 61.               |
| V3                                                                                                                           | AAAACCACGCACACACTCACTCACCG         | 62.               |

|     |                             |     |
|-----|-----------------------------|-----|
| V4  | AAAACGGCCTTCTCCCCGCTCCAACG  | 63. |
| V5  | AAAACGGAGCAAACCTCCCCCACC CG | 64. |
| V6  | AAAACCCACCCCCCTTCCAAAGCCG   | 65. |
| V7  | AAAACCCCCGCCGGGACCCCGCCCCG  | 66. |
| V8  | AAAACAGCGCCACGACCTCCGAGCG   | 67. |
| V9  | AAAACAGCGCCCTGGCCGGTGAGTCG  | 68. |
| V10 | AAAACGTCCACCTAACCGCTGCGCCG  | 69. |
| V11 | AAAACGCGGGGCGGAGCCATGCGCCG  | 70. |
| V12 | AAAACGGTAGCTCGGAGGTCGTGGCG  | 71. |
| V13 | AAAACAGGGGGAGGGCTCACGCCGCG  | 72. |
| V14 | AAAACACCCCCCTCCACCCCGCCTCCG | 73. |
| V15 | AAAACCCGGAGCGCGGCGTGAGCCCG  | 74. |
| V16 | AAAACGAGGCGGGGTGGAGGGGGTTCG | 75. |
| N1  | AAAACGGCCAGATGGCTCCGCGCTCG  | 76. |
| N2  | AAAACCTAACCCCTTCCGCGCCGCGCG | 77. |
| N3  | AAAACGCCGGCCCCGCGCCGCGCCGCG | 78. |
| N4  | AAAACGCGGGGGAGGGCGGCGCGGCG  | 79. |
| N5  | AAAACGGTTGGCTGGTTATAACCGCG  | 80. |
| N6  | AAAACCTCCGAACAGCTCCGCGCACG  | 81. |

These fragments were ligated into BsmBI-digested plasmid pMLM3636 to yield DNA encoding a chimeric ~102 nt single-chain guide RNA (Mali et al., Science. 2013 Feb 15;339(6121):823-6; Hwang et al., Nat Biotechnol. 2013 Mar;31(3):227-9) expressed by a human U6 promoter. The pMLM3636 plasmid and its full DNA sequence are available from Addgene. See Fig. 4.

To engineer a Cas9-activator the D10A, H840A catalytic mutations (previously described in Jinek et al., 2012; and Qi et al., 2013) were introduced into either the wild-type or a codon-optimized Cas9 sequence (Fig. 5). These mutations render the Cas9 catalytically inactive so that it will no longer induce double-strand breaks. In one construct, a triple flag tag, nuclear localization signal and the VP64 activation domain were fused to the C-terminus of the inactive Cas9 (Fig. 6). Expression of this fusion protein was driven by the CMV promoter.

Construction of dCas-VP64 expression plasmids was performed as follows. DNA encoding the Cas9 nuclease harboring inactivating D10A/H840A mutations (dCas9) was amplified by PCR from plasmid pMJ841 (Addgene plasmid #39318) using primers that add a T7 promoter site 5' to the start codon and a nuclear localization signal at the carboxy-terminal end of the Cas9 coding sequences and cloned into a plasmid containing a CMV promoter as previously described (Hwang et al., Nat Biotechnol 31, 227–229 (2013)) to yield plasmid pMLM3629. Oligonucleotides encoding a triple FLAG epitope were annealed and cloned into XhoI and PstI sites in plasmid pMLM3629 to generate plasmid pMLM3647 expressing dCas9 with a C-terminal flag FLAG tag. DNA sequence encoding a Gly<sub>4</sub>Ser linker followed by the synthetic VP64 activation domain was introduced

downstream of the FLAG-tagged dCas9 in plasmid pMLM3647 to yield plasmid pSL690. The D10A/H840A mutations were also introduced by QuikChange site-directed mutagenesis (Agilent) into plasmid pJDS247, which encodes a FLAG-tagged Cas9 sequence that has been codon optimized for expression in human cells, to yield plasmid pMLM3668. DNA sequence encoding the Gly<sub>4</sub>Ser linker and the VP64 activation domain were then cloned into pMLM3668 to yield a codon-optimized dCas9-VP64 expression vector named pMLM3705.

Cell Culture, Transfection and ELISA Assays were performed as follows. Flp-In T-Rex 293 cells were maintained in Advanced DMEM supplemented with 10% FBS, 1% penstrep and 1% Glutamax (Invitrogen). Cells were transfected by Lipofectamine LTX (Invitrogen) according to manufacturer's instructions. Briefly, 160,000 293 cells were seeded in 24-well plates and transfected the following day with 250ng gRNA plasmid, 250ng Cas9-VP64 plasmid, 30ng pmaxGFP plasmid (Lonza), 0.5ul Plus Reagent and 1.65ul Lipofectamine LTX. Tissue culture media from transfected 293 cells was harvested 40 hours after transfection, and secreted VEGF-A protein assayed using R&D System's Human VEGF-A ELISA kit "Human VEGF Immunoassay."

16 sgRNAs were constructed for target sequences within three DNase I hypersensitive sites (HSSs) located upstream, downstream or at the transcription start site of the human *VEGFA* gene in 293 cells (**Fig. 1E**).

Before testing the abilities of the 16 *VEGFA*-targeted gRNAs to recruit a novel dCas9-VP64 fusion protein, each of these gRNAs was first assessed for its ability to direct Cas9 nuclease to its intended target site in human 293 cells. For this purpose, gRNA and Cas9 expression vectors were transfected in a 1:3 ratio because previous optimization experiments demonstrated a high level of Cas9-induced DNA cleavage in U2OS cells using this ratio of plasmids.

Transfections of 293 cells were performed as described above for the dCas9-VP16 *VEGFA* experiments except that cells were transfected with 125 ng of plasmid encoding *VEGFA*-targeted gRNAs and 375 ng of plasmid encoding active Cas9 nuclease (pMLM3639). 40 hours post-transfection, genomic DNA was isolated using the QIAamp DNA Blood Mini kit (Qiagen) according to manufacturer's instructions. PCR amplification of the three different targeted regions in the *VEGFA* promoter was performed using Phusion Hot Start II high-fidelity DNA polymerase (NEB) with 3% DMSO and the following touchdown PCR cycle: 10 cycles of 98 °C, 10 s; 72–62 °C, –1 °C/cycle, 15 s; 72 °C, 30 s, followed by 25 cycles of 98 °C, 10 s; 62 °C, 15 s; 72 °C, 30 s. The -500 region was

amplified using primers oFYF434 (5'- TCCAGATGGCACATTGTCAG-3' (SEQ ID NO:82)) and oFYF435 (5'- AGGGAGCAGGAAAGTGAGGT-3' (SEQ ID NO:83)). The region around the transcription start site was amplified using primers oFYF438 (5'- GCACGTAACCTCACTTTCCT-3' (SEQ ID NO:84)) and oFYF439 (5'-

CTTGCTACCTCTTTCCTCTTCT-3' (SEQ ID NO:85)). The +500 region was amplified using primers oFYF444 (5'- AGAGAAGTCGAGGAAGAGAGAG-3' (SEQ ID NO:86)) and oFYF445 (5'- CAGCAGAAAGTTCATGGTTTCG-3' (SEQ ID NO:87)). PCR products were purified using Ampure XP beads (Agencourt) and T7 Endonuclease I assays were performed and analyzed on a QIAXCEL capillary electrophoresis system as previously described (Reyon et al., Nat Biotech 30, 460-465 (2012)).

All 16 gRNAs were able to mediate the efficient introduction of Cas9 nuclease-induced indel mutations at their respective target sites as assessed using a previously described T7E1 genotyping assay (Table 2). Thus all 16 gRNAs can complex with Cas9 nuclease and direct its activity to specific target genomic sites in human cells.

**Table 2. Frequencies of indel mutations induced by VEGFA-targeted gRNAs and Cas9 nuclease**

| gRNA | Mean Indel Mutation Frequency (%) $\pm$ SEM |
|------|---------------------------------------------|
| V1   | 18.05 $\pm$ 0.47                            |
| V2   | 41.48 $\pm$ 0.62                            |
| V3   | 33.22 $\pm$ 1.05                            |
| V4   | 16.97 $\pm$ 0.06                            |
| V5   | 7.46 $\pm$ 0.50                             |
| V6   | 16.99 $\pm$ 0.51                            |
| V7   | 1.42 $\pm$ 0.11                             |
| V8   | 34.07 $\pm$ 0.90                            |
| V9   | 24.53 $\pm$ 1.40                            |
| V10  | 35.65 $\pm$ 1.35                            |
| V11  | 4.45 $\pm$ 0.22                             |
| V12  | 23.95 $\pm$ 0.41                            |
| V13  | 9.45 $\pm$ 0.74                             |
| V14  | 12.17 $\pm$ 0.36                            |
| V15  | 14.28 $\pm$ 0.54                            |
| V16  | 18.82 $\pm$ 1.48                            |

To test whether dCas9-VP64 protein could also be targeted to specific genomic sites in human cells by these same gRNAs, Enzyme-Linked Immunoblot Assays of VEGFA protein were performed as follows. Culture medium of Flp-In T-Rex HEK293 cells transfected with plasmids encoding VEGFA-targeted sgRNA and dCas9-VP64 was harvested 40 hours post-transfection and VEGFA protein expression was measured by

ELISA as previously described (Maeder et al., Nat Methods 10, 243–245 (2013)). Fold-activation of VEGFA expression was calculated by dividing the concentration of VEGFA protein in media from cells in which both a sgRNA and dCas9-VP64 were expressed by the concentration of VEGFA protein in media from cells in which an off-target sgRNA

(targeted to a sequence in the *EGFP* reporter gene) and dCas9-VP64 were expressed.

15 of the 16 gRNAs tested induced significant increases in VEGFA protein expression when co-expressed with dCas9-VP64 in human 293 cells (Fig. 2A). The magnitude of VEGFA induction observed ranged from two- to 18.7-fold-activation with a mean of five-fold-activation. Control experiments revealed that expression of each of the 16 gRNAs alone, dCas9-VP64 alone, and dCas9-VP64 together with an “off-target” gRNA designed to bind an EGFP reporter gene sequence all failed to induce elevated VEGFA expression (Fig. 2A), demonstrating that co-expression of a specific gRNA and the dCas9-VP64 protein are both required for promoter activation. Thus dCas9-VP64 is stably expressed and can be directed by gRNAs to activate transcription of specific genomic loci in human cells. The greatest increase in VEGFA was observed in cells transfected with gRNA3, which induced protein expression by 18.7-fold. Interestingly, the three best gRNAs, and 6 of the 9 gRNAs capable of inducing expression by 3-fold or more, target the -500 region (~500bp upstream of the transcription start site).

Because in one aspect the system described herein uses variable gRNAs to recruit a common dCas9-VP64 activator fusion, one can envision that the expression of multiple guide RNAs in a single cell might enable multiplex or combinatorial activation of endogenous gene targets. To test this possibility, 293 cells were transfected with dCas9-VP64 expression plasmid together with expression plasmids for four gRNAs (V1, V2, V3, and V4) that each individually induced expression from the VEGFA promoter. Co-expression of all four gRNAs with dCas9-VP64 induced synergistic activation of VEGFA protein expression (i.e., a fold-activation greater than the expected additive effects of each individual activator) (Fig. 2B). In addition, various combinations of three of these four activators also activated the VEGFA promoter synergistically (Fig. 2B). Because synergistic activation of transcription is believed to result from the recruitment of multiple activator domains to a single promoter, multiple gRNA/dCas9-VP64 complexes are likely to be simultaneously binding to the VEGFA promoter in these experiments.

These experiments demonstrate that co-expression of a Cas9-HFD, e.g., a Cas9-activator protein (harboring the VP64 transcriptional activation domain) and a sgRNA with 20nt of sequence complementarity to sites in the human VEGF-A promoter in human

HEK293 cells can result in upregulation of VEGF-A expression. Increases in VEGF-A protein were measured by ELISA assay and it was found that individual gRNAs can function together with a Cas9-activator fusion protein to increase VEGF-A protein levels by up to ~18-fold (Fig. 2A). Additionally, it was possible to achieve even greater increases in activation through transcriptional synergy by introducing multiple gRNAs targeting various sites in the same promoter together with Cas9-activator fusion proteins (Fig. 2B).

### **Example 2. Engineering CRISPR/Cas Activator System targeting the endogenous human *NTF3* gene**

To extend the generality of the present findings, we tested whether the RNA-guided activator platform could be used to induce the expression of the human *NTF3* gene. To do this, six sgRNAs were designed to a predicted DNase I hypersensitive site (HSS) in the human *NTF3* promoter and plasmids expressing each of these gRNAs were co-transfected with a plasmid encoding dCas9-VP64 protein that had been codon optimized for human cell expression (Fig. 3A).

All six gRNAs tested induced significant increases in *NTF3* transcript levels as detected by quantitative RT-PCR (Fig. 3B). Although fold-activation values for these six RNA-guided activators could not be accurately calculated (because basal levels of transcript were essentially undetectable), the mean levels of activated *NTF3* mRNA expression varied over a four-fold range. Decreasing the amounts of gRNA and dCas9-VP64 expression plasmids transfected resulted in less activation of the *NTF3* gene (Fig. 3B), demonstrating a clear dose-dependent effect.

In addition, 293 cells were co-transfected with dCas9-VP64 and *NTF3*-targeted gRNA expression plasmids alone and in single and double combinations. Relative expression of *NTF3* mRNA was detected by quantitative RT-PCR and normalized to a GAPDH control ( $\Delta\text{Ct} \times 10^4$ ). In all of these experiments the amount of each individual gRNA expression plasmid used for transfection was the same. FIG. 3B shows that this multiplex gRNA expression induced synergistic activation of *NTF3* mRNA expression by dCas9-VP64 protein.

### **Example 3. Engineering CRISPR/Cas-MS2, -Csy4 and -Lambda N Fusion Systems – Creating Biological Tethers**

Fusion proteins are made in which an MS2 coat protein, Csy4 nuclease (preferably catalytically inactive Csy4, e.g., the H29A mutant described in Haurwitz et al.

329(5997):1355-8 (2010)), or the lambda N are fused to the N- or C-terminus of the inactivated dCas9. MS2 and lambda N are bacteriophage proteins that bind to a specific RNA sequence, and thus can be used as adapters to tether to the dCas9 protein a heterologous RNA sequence tagged with the specific MS2 or lambda N RNA binding sequence. dCas9-MS2 fusions or dCas9-lambda N fusions are co-expressed with chimeric long non-coding RNAs (lncRNAs) fused to the MS2 or lambda N stem loop recognition sequence on either their 5' or 3' end. Chimeric Xist or chimeric RepA lncRNAs will be specifically recruited by the dCas9 fusions and the ability of this strategy to induce targeted silencing will be assayed by measuring target gene expression. The system will be optimized by testing various alterations to the coat proteins and chimeric RNAs. The N55K and deltaFG mutations to the MS2 coat protein have been previously demonstrated to prevent protein aggregation and increase affinity for the stem-loop RNA. Additionally, we will test the high-affinity C-loop RNA mutant reported to increase affinity for the MS2 coat protein. Exemplary sequences for the MS2 and lambda N proteins are given below; the MS2 functions as a dimer, therefore the MS2 protein can include a fused single chain dimer sequence.

**1. Exemplary sequences for Fusions of single MS2 coat protein (wt, N55K or deltaFG) to the N-terminus or C-terminus of the dCas9.**

MS2 coat protein amino acid sequence:

MASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQ  
NRKYTIKVEVPKVATQTVGGVELPVAAWRSYLNMEITIPFATNSDCELVKAMQGLKDG  
NPIPSAIAANSGLY (SEQ ID NO:88)

MS2 N55K:

MASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQ  
KRKYTIKVEVPKVATQTVGGVELPVAAWRSYLNMEITIPFATNSDCELVKAMQGLKDG  
NPIPSAIAANSGLY (SEQ ID NO:89)

MS2deltaFG:

MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQ  
NRKYTIKVEVPKGAWRSYLNMEITIPFATNSDCELVKAMQGLKDG  
NPIPSAIAANSGLY (SEQ ID NO:90)

**2. Exemplary sequences for Fusions of fused dimeric MS2 coat protein (wt, N55K or deltaFG) to the N-terminus or C-terminus of dCas9.**

Dimeric MS2 coat protein:

MASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQ  
NRKYTIKVEVPKVATQTVGGVELPVAAWRSYLNMEITIPFATNSDCELVKAMQGLKDG  
NPIPSAIAANSGLYMASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKVATQTVGGVELPVAAWRSYLNMEITIPFATNSDCELVKAMQGLKDG  
NPIPSAIAANSGLY (SEQ ID NO:91)

MASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQ  
 KRKYTIKVEVPKVATQTVGGVELPVAAWRSYLNMEITPIFATNSDCELVKAMQG  
 LLKDGNIPIPSAIAANSGLYGAMASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWI  
 SSNSRSQAYKVTCSVRQSSAQKRKYTIKVEVPKVATQTVGGVELPVAAWRSYLN  
 5 MELTPIFATNSDCELVKAMQGLLKDGNIPIPSAIAANSLIN (SEQ ID NO:92)

Dimeric MS2deltaFG:

MASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQ  
 KRKYTIKVEVPKGAWRSYLNMEITPIFATNSDCELVKAMQGLLKDGNIPIPSAIAA  
 NSGLYGAMASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCS  
 10 VRQSSAQKRKYTIKVEVPKGAWRSYLNMEITPIFATNSDCELVKAMQGLLKDGNI  
 PIPSAIAANSLIN (SEQ ID NO:93)

### 3. Exemplary sequences for Fusions of Lambda N to N-terminus or C-terminus of dCas9.

Lambda N amino acid sequence:

MDAQTRRRERRAEKQAQWKAAN (SEQ ID NO:94) or  
 MDAQTRRRERRAEKQAQWKAANPLLVGVSAPVNRPIPSLNRPKPKSRVESALNPI  
 DLTVLAIEYHKQIESNLQRIERKNQRTWYSKPGERGITCSGRQKIKGKSIPLI (SEQ  
 ID NO:95)

### 4. Exemplary sequence for Fusions of Csy4 to N-terminus or C-terminus of dCas9

Exemplary sequences for Cys4 are given in Haurwitz et al. 329(5997):1355-8  
 (2010), e.g., the inactivated form.

The constructs are expressed in cells also expressing a regulatory RNA, e.g., a long  
 non-coding RNA (lncRNA) such as HOTAIR, HOTTIP, XIST or XIST RepA, that has  
 25 been fused with the cognate stem-loop recognition sequence for the lambda N or MS2 on  
 either its 5' or 3' end. The wild type and high-affinity sequences for MS2 are  
 AAACAUGAGGAUUACCCAUGUCG (SEQ ID NO:96) and  
 AAACAUGAGGAUCACCCAUGUCG (SEQ ID NO:97), respectively (see Keryer-Bibens  
 et al., supra, FIG. 2); the nutL and nutR BoxB sequences to which lambda N binds are  
 30 GCCCUGAAGAAGGGC (SEQ ID NO:98) and GCCCUGAAAAAGGGC (SEQ ID  
 NO:99), respectively. The sequence to which Csy4 binds is  
 GTTCACTGCCGTATAGGCAG (truncated 20 nt) (SEQ ID NO:100) or  
 GUUCACUGCCGUAUAGGCAGCUAAGAAA (SEQ ID NO:101).

The binding of the dCas9/MS2 to a target site in a cell expressing an MS2-binding  
 35 sequence tagged lncRNA recruits that lncRNA to the dCas9 binding site; where the  
 lncRNA is a repressor, e.g., XIST, genes near the dCas9 binding site are repressed.  
 Similarly, binding of the dCas9/lambdaN to a target site in a cell expressing an lambdaN-  
 binding sequence tagged lncRNA recruits that lncRNA to the dCas9 binding site.

#### **Example 4. Engineering CRISPR/Cas-HP1 Fusion Systems –Sequence-Specific Silencing**

The dCas9 fusion proteins described herein can also be used to target silencing domains, e.g., Heterochromatin Protein 1 (HP1, also known as swi6), e.g., HP1 $\alpha$  or HP1 $\beta$ .

5 Truncated versions of HP1 $\alpha$  or HP1 $\beta$  in which the chromodomain has been removed can be targeted to specific loci to induce heterochromatin formation and gene silencing.

Exemplary sequences of truncated HP1 fused to dCas9 are shown in Figs. 8A-8B. The HP1 sequences can be fused to the N- or C-terminus of the inactivated dCas9 as described above.

#### **Example 5. Engineering CRISPR/Cas-TET Fusion Systems –Sequence-Specific Demethylation**

The dCas9 fusion proteins described herein can also be used to target enzymes that modify the methylation state of DNA (e.g., DNA methyltransferase (DNMT) or TET proteins). Truncated versions of TET1 can be targeted to specific loci to catalyze DNA  
15 demethylation. Exemplary sequences of truncated TET1 fused to dCas9 are shown in Fig. 9. The TET1 sequence can be fused to the N- or C-terminus of the inactivated dCas9 as described above.

#### **Example 6. Engineering Optimized CRISPR/Cas-VP64 Fusions**

The activities of dCas9-based transcription activators harboring the VP64 activation  
20 domain were optimized by varying the number and position of the nuclear localization signal(s) (NLS) and 3xFLAG-tags within these fusions (Figure 10). dCas9-VP64 fusions that contain both an N-terminal NLS and an NLS that lies between the dCas9 and VP64 sequences consistently induce higher levels of target gene activation, perhaps resulting from enhanced nuclear localization of the activator (Figure 10). Furthermore, even greater  
25 levels of activation were observed when a 3xFLAG tag was placed between the C-terminal end of dCas9 and the N-terminal end of VP64. The 3xFLAG tag may act as an artificial linker, providing necessary spacing between dCas9 and VP64 and perhaps allowing for better folding of the VP64 domain (that may not be possible when constrained near dCas9) or better recognition of VP64 by transcriptional mediator complexes that recruit RNA  
30 polymerase II. Alternatively, the negatively charged 3xFLAG tag might also function as a fortuitous transcriptional activation domain, enhancing the effects of the VP64 domain.

**Example 7. Optimized Catalytically Inactive Cas9 Proteins (dCas9)**

Additional optimization of the activities of dCas9-VP64 activators was performed by changing the nature of the inactivating mutations that abolish the nuclease activity of Cas9 in the dCas9 domain (Figure 11A-B). In published studies to date, the catalytic residues D10 and H840 were mutated to alanine (D10A and H840A) to disrupt the active site networks that mediate the hydrolysis of DNA. It was hypothesized that alanine substitutions at these positions might result in destabilization of dCas9 and therefore suboptimal activity. Therefore, more structurally conservative substitutions at D10 or H840 (for example, to asparagine or tyrosine residues: D10N, H840N, and H840Y) were tested to see if they might lead to greater gene activation by dCas9-VP64 fusions bearing these different mutations. When dCas9-VP64 variants bearing these variant substitutions were co-transfected into HEK293 cells with three gRNAs targeting upstream regions of the endogenous human VEGFA gene, greater VEGFA protein expression was observed for all but one of these variants (Figure 11A). However, this effect was not as significant when the dCas9-VP64 variants were co-transfected with only one of these gRNAs (Figure 11A), or when transfected into a HEK293 derivative cell-line that expresses a single VEGFA-targeted gRNA (Figure 11B).

**OTHER EMBODIMENTS**

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

## WHAT IS CLAIMED IS:

1. A fusion protein comprising catalytically inactive CRISPR associated 9 (dCas9) protein linked to a heterologous functional domain.
2. The fusion protein of claim 1, wherein the heterologous functional domain is a transcriptional activation domain.
3. The fusion protein of claim 2, wherein the transcriptional activation domain is from VP64 or NF- $\kappa$ B p65.
4. The fusion protein of claim 1, wherein the heterologous functional domain is a transcriptional silencer or transcriptional repression domain.
5. The fusion protein of claim 4, wherein the transcriptional repression domain is a Krueppel-associated box (KRAB) domain, ERF repressor domain (ERD), or mSin3A interaction domain (SID).
6. The fusion protein of claim 4, wherein the transcriptional silencer is Heterochromatin Protein 1 (HP1), e.g., HP1 $\alpha$  or HP1 $\beta$ .
7. The fusion protein of claim 1, wherein the heterologous functional domain is an enzyme that modifies the methylation state of DNA.
8. The fusion protein of claim 7, wherein the enzyme that modifies the methylation state of DNA is a DNA methyltransferase (DNMT) or a TET protein.
9. The fusion protein of claim 8, wherein the TET protein is TET1.
10. The fusion protein of claim 1, wherein the heterologous functional domain is an enzyme that modifies a histone subunit.
11. The fusion protein of claim 1, wherein the enzyme that modifies a histone subunit is a histone acetyltransferase (HAT), histone deacetylase (HDAC), histone methyltransferase (HMT), or histone demethylase.

12. The fusion protein of claim 1, wherein the heterologous functional domain is a biological tether.
13. The fusion protein of claim 12, wherein the biological tether is MS2, Csy4 or lambda N protein.
14. The fusion protein of claim 1, wherein the catalytically inactive Cas9 protein is from *S. pyogenes*.
15. The fusion protein of claim 1, wherein the catalytically inactive Cas9 protein comprises mutations at D10, E762, H983, or D986; and at H840 or N863.
16. The fusion protein of claim 15, wherein the mutations are:
  - (i) D10A or D10N, and
  - (ii) H840A, H840N, or H840Y.
17. The fusion protein of claim 1, wherein the heterologous functional domain is linked to the N terminus or C terminus of the catalytically inactive Cas9 protein, with an optional intervening linker, wherein the linker does not interfere with activity of the fusion protein.
18. The fusion protein of claim 1, further comprising one or both of a nuclear localization sequence and one or more epitope tags on the N-terminus, C-terminus, and/or in between the catalytically inactive CRISPR associated 9 (Cas9) protein and the heterologous functional domain, optionally with one or more intervening linkers.
19. The fusion protein of claim 18, wherein the epitope tag is c-myc, 6His, or FLAG.
20. A nucleic acid encoding the fusion protein of any of claims 1-19.
21. An expression vector comprising the nucleic acid of claim 20.
22. A method of increasing expression of a target gene in a cell, the method comprising expressing the fusion protein of claims 2-3 in the cell, and one or more guide RNAs directed to the target gene.

23. A method of decreasing of a target gene in a cell, the method comprising expressing the fusion protein of claims 4-6 in the cell, and one or more guide RNAs directed to the target gene.
24. A method of reducing DNA methylation of a target gene or its promoter or enhancer(s) in a cell, the method comprising expressing the fusion protein of claims 7-9 in the cell, and one or more guide RNAs directed to the relevant target gene sequence.
25. A method of modifying histones associated with a target gene or its promoter or enhancer(s) in a cell, the method comprising expressing the fusion protein of claims 10-11 in the cell, and one or more guide RNAs directed to the relevant target gene sequence.

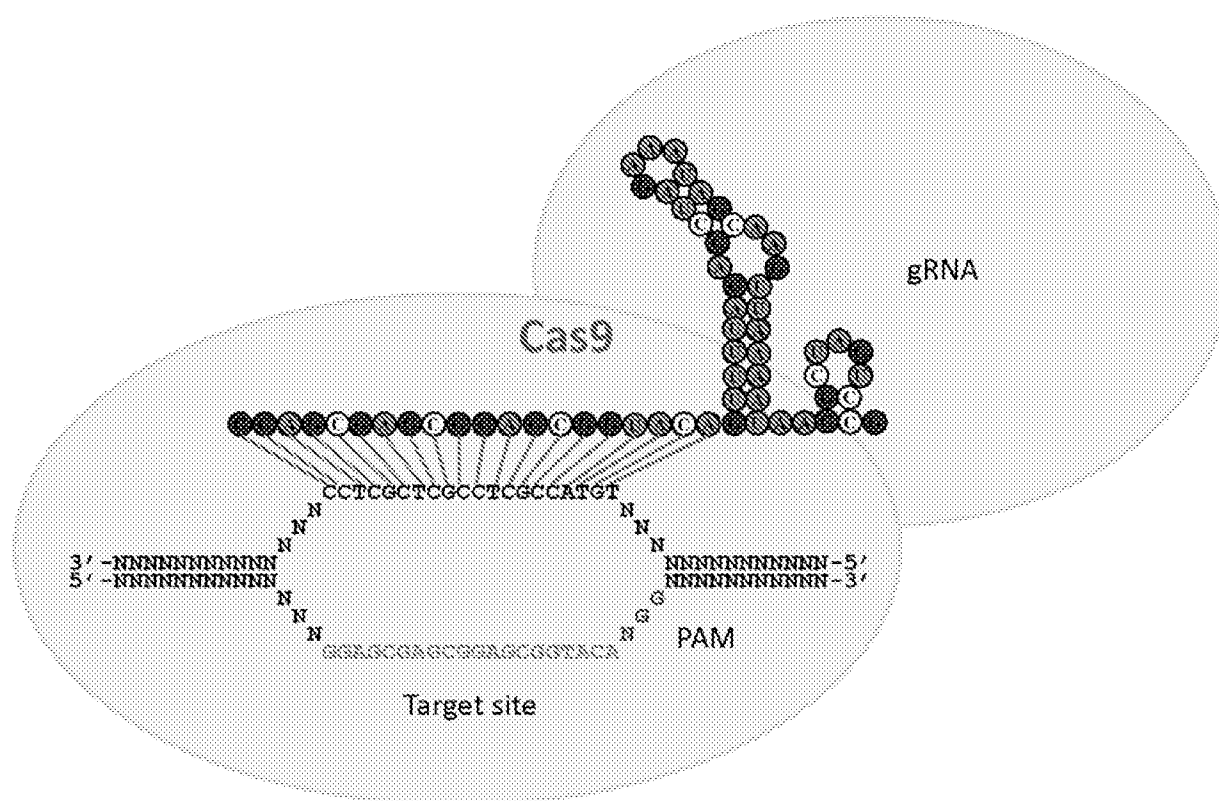


FIG. 1A

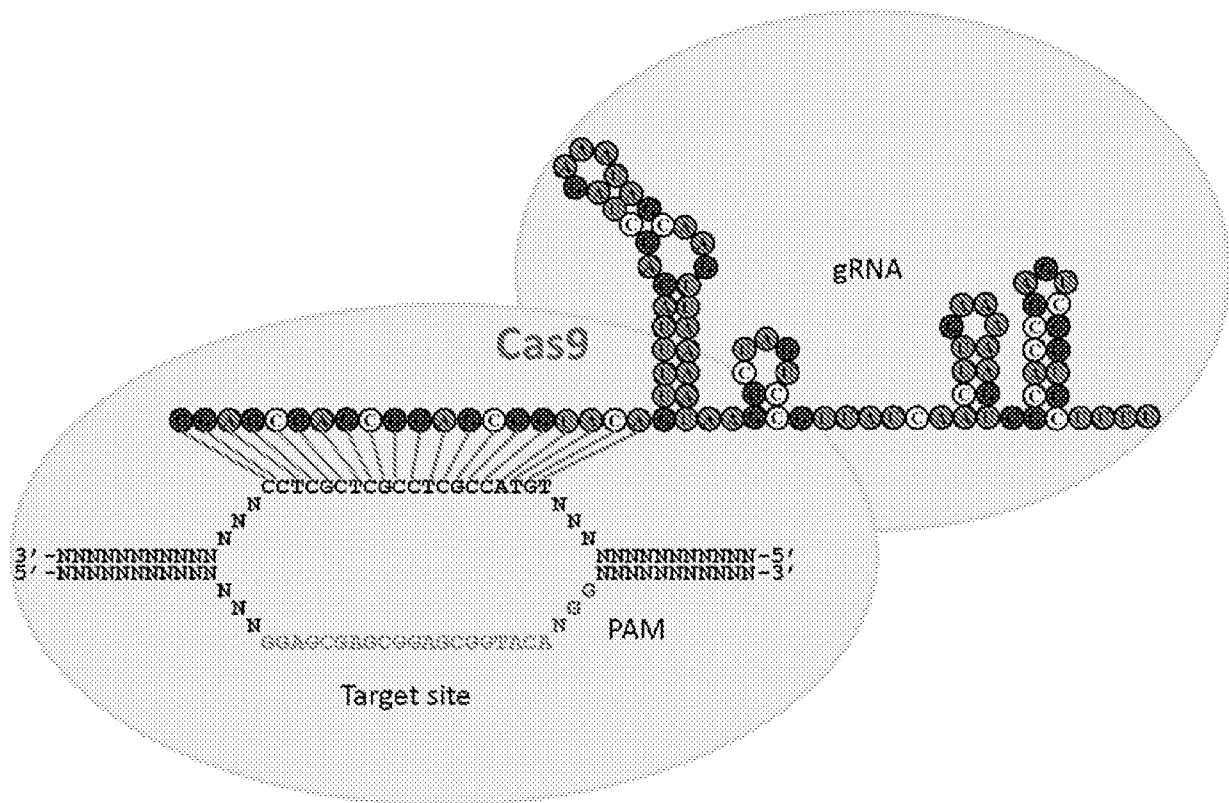


FIG. 1B

3 / 20

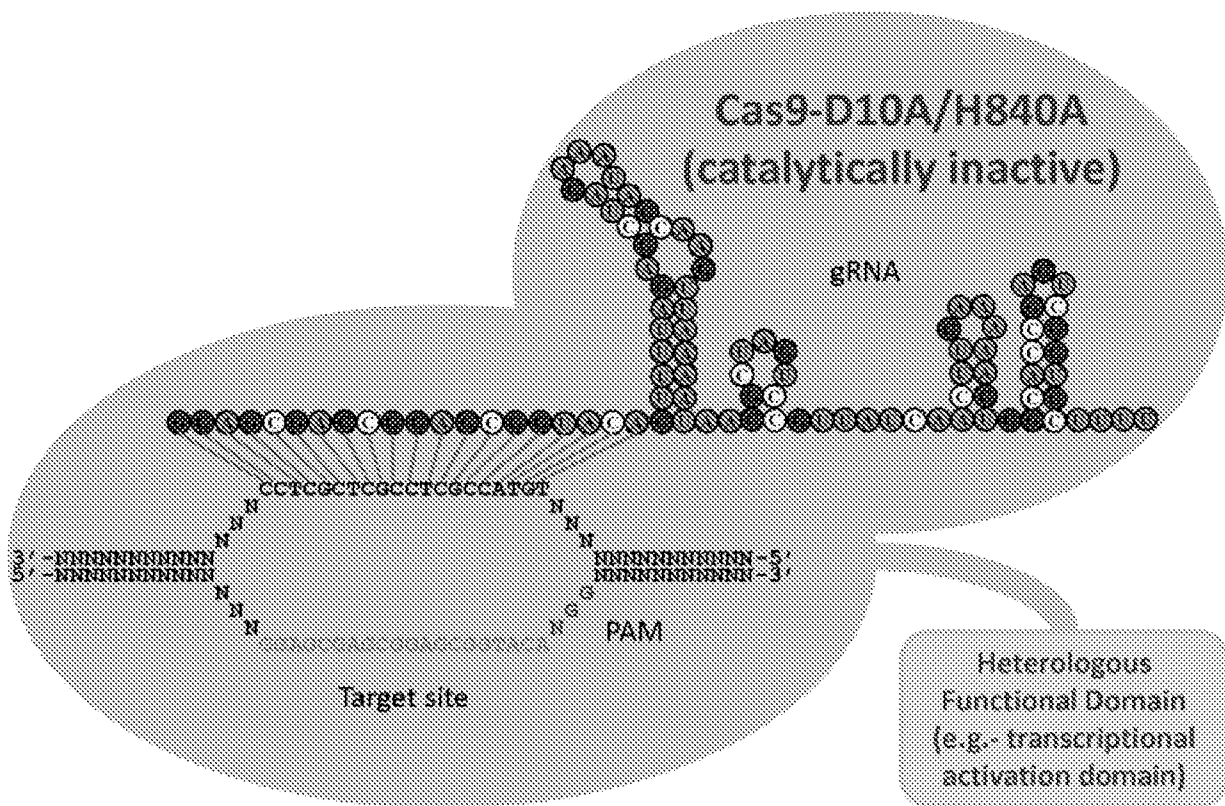


FIG. 1C

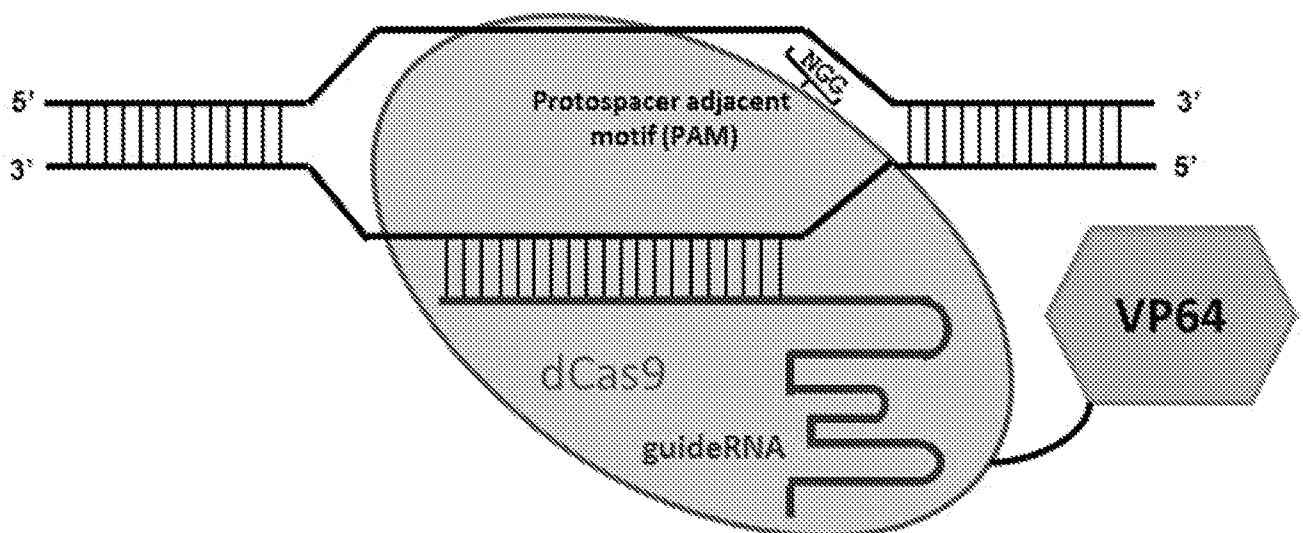


FIG. 1D

4 / 20

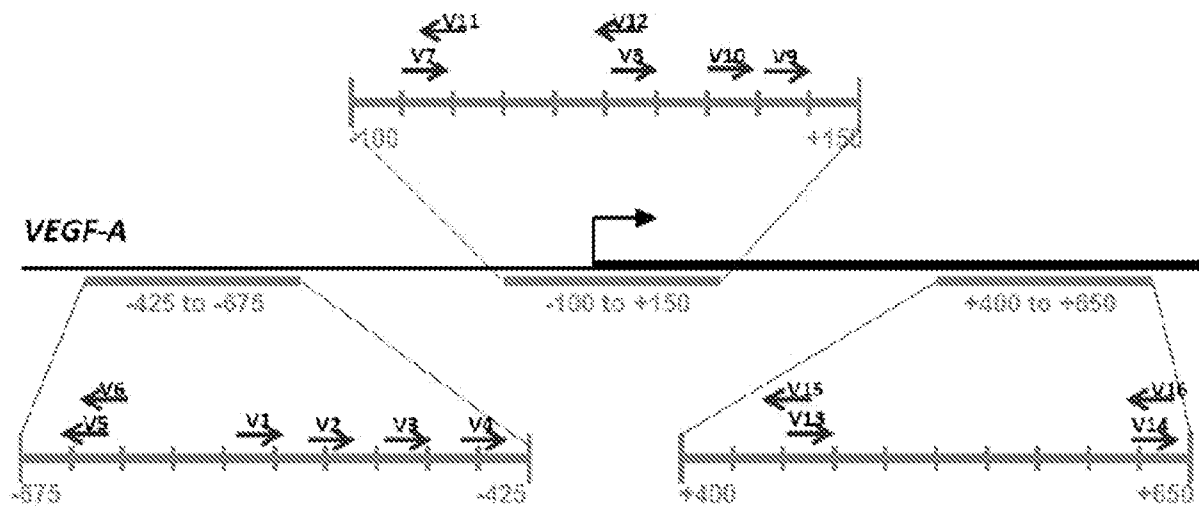


FIG. 1E

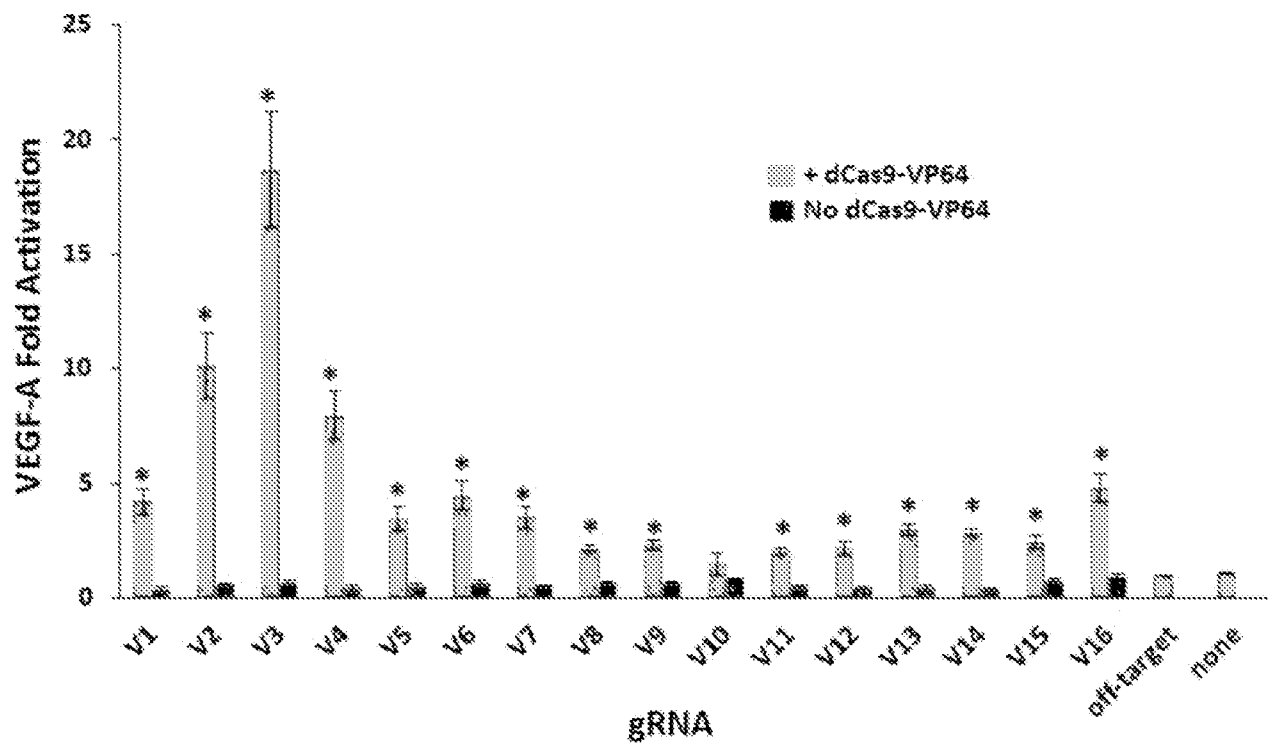


FIG. 2A

5 / 20

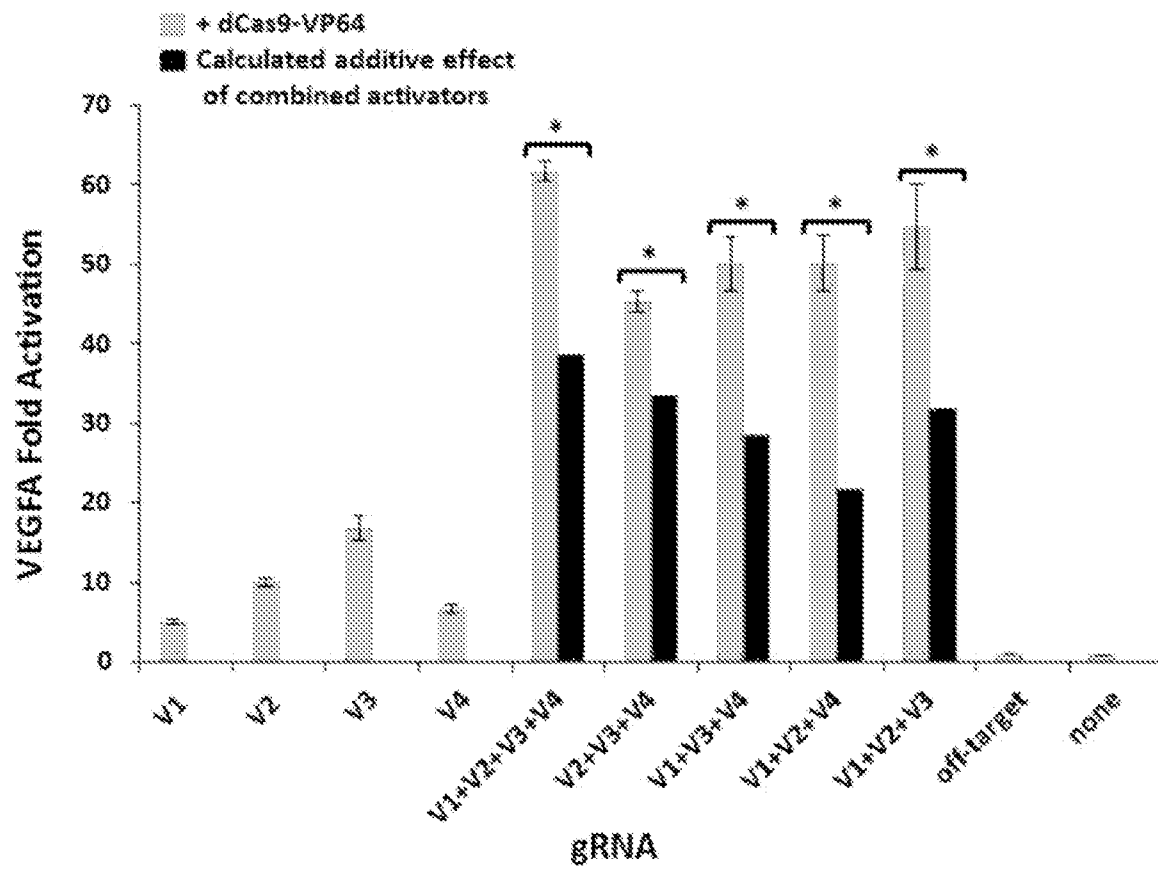


FIG. 2B

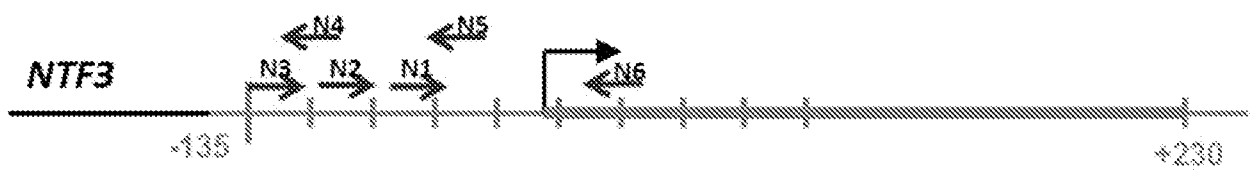


FIG. 3A

6 / 20

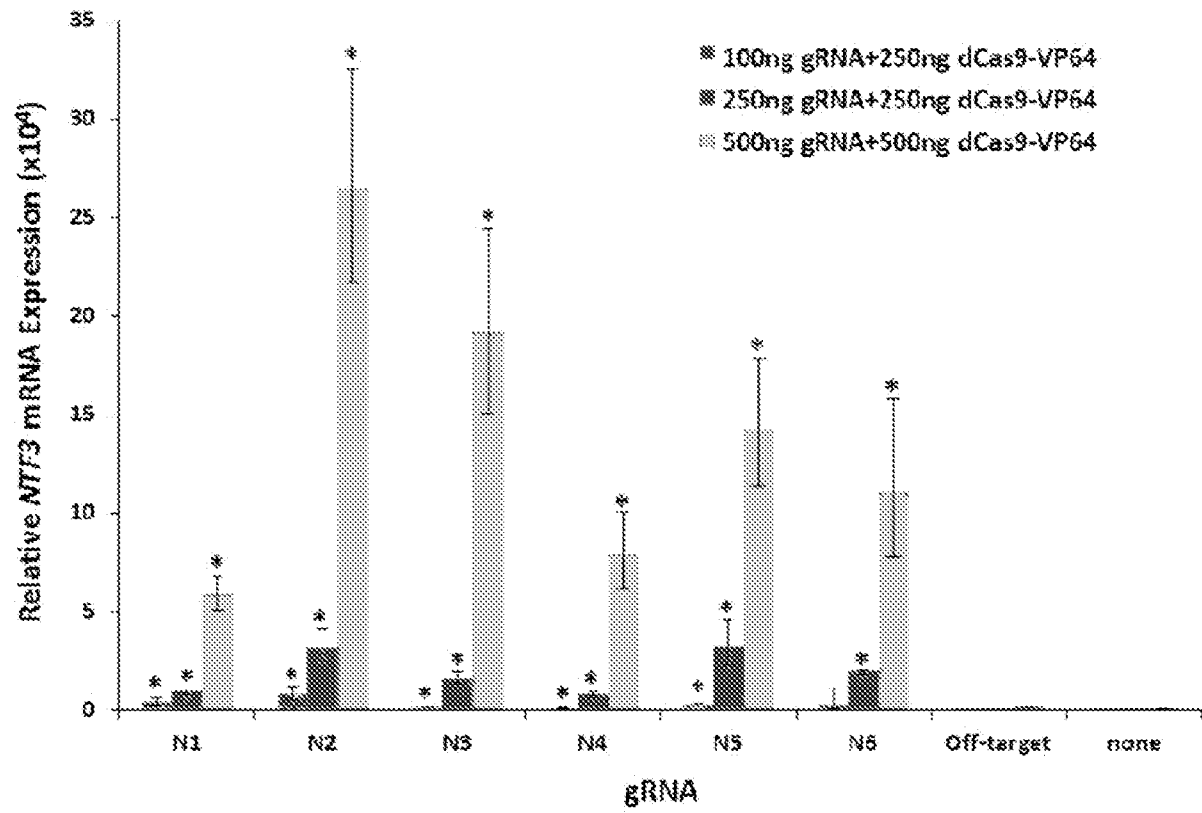


FIG. 3B

7 / 20

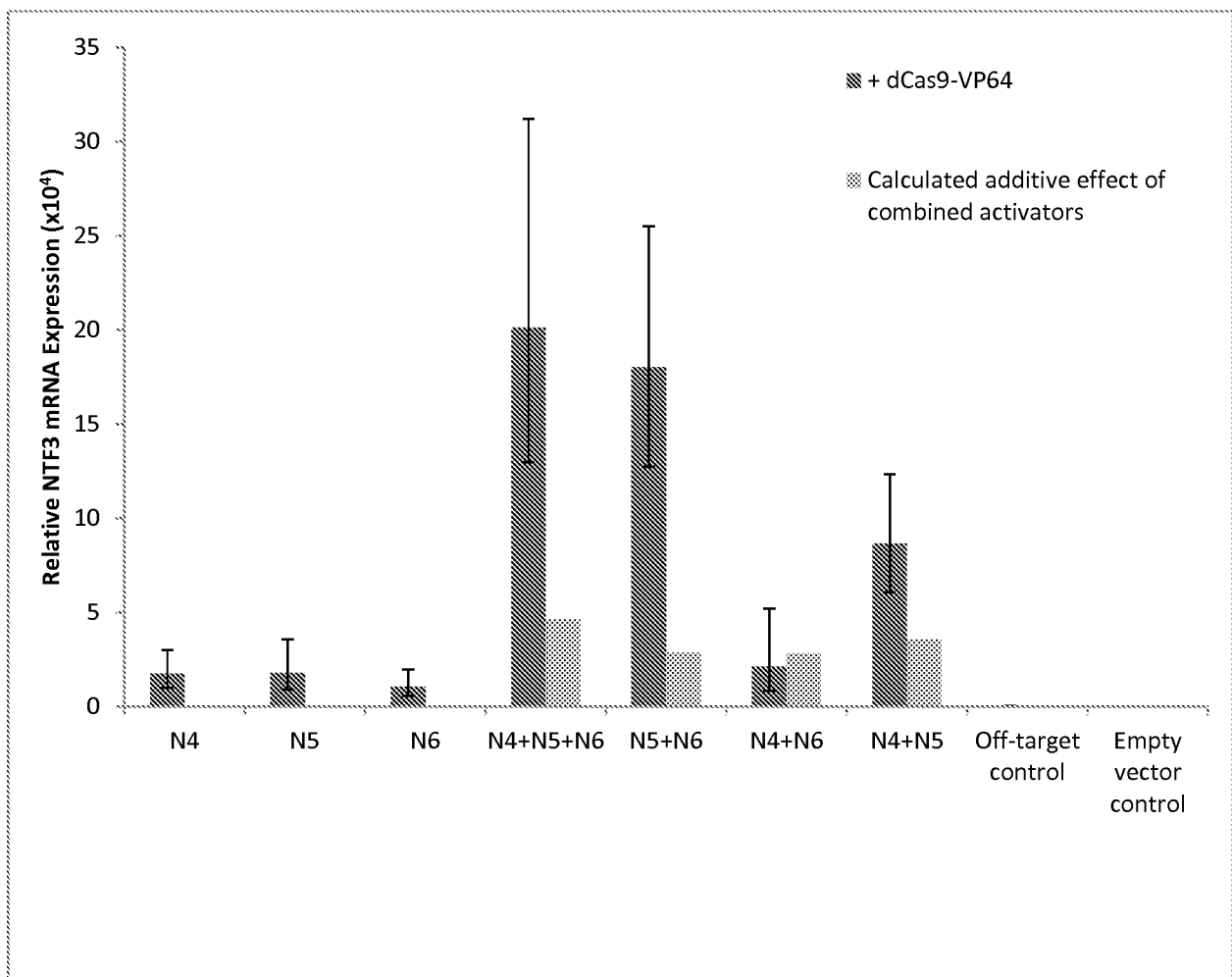


FIG. 3C

**FIG. 4 - Guide RNA expression vector sequence**

GACGTCGCTAGCTGTACAAAAAAGCAGGCTTTAAAGGAACCAATTCAGTCGACTGGATCCGGTACCAA  
GGTCGGGCAGGAAGAGGGCCTATTTCCCATGATTCCCTTCATATTTGCATATACGATACAAGGCTGTTA  
GAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAAG  
TAATAATTTCTTGGGTAGTTTGCAGTTTTAAATATATGTTTTAAATGGACTATCATATGCTTACCGT  
AACTTGAAAGTATTTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACNNNNNNNNNN  
NNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAA  
GTGGCACCGAGTCGGTGCTTTTTTTAAGCTTGGGCCGCTCGAGGTACCTCTCTACATATGACATGTGA  
GCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCG  
CCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAA  
GATACCAGGCGTTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGA  
TACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAG  
TTCGGTGTAGGTCGTTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCCTGCG  
CCTTATCCGGTAACATATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCC  
ACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAA  
CTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAA  
GAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAG  
CAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCA  
GTGGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCC  
TTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAACTTGGTCTGACAGTTAC  
CAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACT  
CCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGC  
GAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGA  
AGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGAAGCTAGAGTAAGTAG  
TTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGT  
TTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGC  
AAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACT  
CATGGTTATGGCAGCACTGCATAATTCTCTTACTGTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTG  
GTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCA  
ATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGG  
GCGAAAACCTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGACCCAACT  
GATCTTCAGCATCTTTTACTTTACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCA  
AAAAAGGGAATAAGGGCGACACGGAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAG  
CATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAACAAATAG  
GGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCT (SEQ ID NO:107)

FIG. 5 - CMV-T7-Cas9 D10A/H840A-3XFLAG-VP64:

ATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGAC  
CTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGTATGCGGTTTTGGC  
AGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCTAAGTCTCCACCCATTGACGTCAATGG  
GAGTTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACCTCCGCCCCATTGACGCAAATG  
GGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGAG  
ATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACCATGGATAAGAAATACTCAATAGGCTTAGc  
TATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGATGAATATAAGGTTCCGTCTAAAAAGTTCAAGGTT  
CTGGGAAATACAGACCGCCACAGTATCAAAAAAATCTTATAGGGGCTCTTTTATTTGACAGTGGAGAGACAG  
CGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACA  
GGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTGTTGG  
AAGAAGACAAGAAGCATGAACGTCATCTATTTTTGGAAATATAGTAGATGAAGTTGCTTATCATGAGAAATA  
TCCAATCTATCATCTGCGAAAAAATTGGTAGATTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGG  
CCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTG  
GACAACTATTTATCCAGTTGGTACAAACCTACAATCAATTATTTGAAGAAAACCTATTAACGCAAGTGGAGT  
AGATGCTAAAGCGATTCTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCG  
GTGAGAAGAAAAATGGCTTATTTGGGAATCTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTTAAATCAAAT  
TTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAGATACTTACGATGATGATTTAGATAATTTATTGGC  
GCAAATTGGAGATCAATATGCTGATTTGTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTAGATAT  
CCTAAGAGTAAATACTGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAACGCTACGATGAACATCATC  
AAGACTTGACTCTTTTAAAGCTTTAGTTTCGACAACAACCTCCAGAAAAGTATAAAGAAATCTTTTTGATCAAT  
CAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAAT  
TTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAACTAAATCGTGAAGATTTGCTGCGCAAGCAACG  
GACCTTTGACAACGGCTCTATTTCCCATCAAATTCATTGGGTGAGCTGCATGCTATTTGAGAAGACAAGAA  
GACTTTTATCCATTTTTTAAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTT  
GGTCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGA  
ATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAACTTTGATAAAAAAT  
CTTCCAAATGAAAAAGTACTACCAAACATAGTTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAA  
GGTCAAATATGTTACTGAAGGAATGCGAAAACCAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGAT  
TACTCTTCAAACAAATCGAAAAAGTAACCGTTAAGCAATTAAGAAGATTATTTCAAAAAAATAGAATGTTT  
TGATAGTGTTGAAATTTAGGAGTTGAAGATAGATTTAATGCTTCATTAGGTACCTACCATGATTTGCTAAAAA  
TTATTAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTAAACATTGACC  
TTATTTGAAGATAGGGAGATGATTGAGGAAAGACTTAAACATATGCTCACCTCTTTGATGATAAGGTGATGA  
AACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGCTCGAAAATTGATTAATGGTATTAGGGATAA  
GCAATCTGGCAAAACAATATTAGATTTTTTGAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCC  
ATGATGATAGTTTGACATTTAAAGAAGACATTCAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTACATGA  
ACATATTGCAATTTAGCTGGTAGCCCTGCTATTAATAAAGGTATTTTACAGACTGTAAAAGTTGTTGATGAAT  
TGGTCAAAGTAATGGGGCGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATCAGACAACCTC  
AAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAG  
ATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTCCAAAATGG  
AAGAGACATGTATGTGGACCAAGAATTAGATATTAATCGTTTAAAGTATTGATGTCGATgcATTGTTCCAC  
AAAGTTTCCTTAAAGACGATTCAATAGACAATAAGGTCTTAACGCGTTCTGATAAAAAATCGTGGTAAATCGGA  
TAACGTTCCAAGTGAAGAAGTAGTCAAAAAAGATGAAAAACTATTGGAGACAACCTCTAAACGCCAAGTTAATC  
ACTCAACGTAAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAACCTTGATAAAGCTGGTTTTA  
TCAAACGCCAATTGGTTGAACTCGCCAAATCACTAAGCATGTGGCACAAATTTGGATAGTCGCATGAATAC  
TAAATACGATGAAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAATCTAAATTAGTTTCTGACT  
TCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAAT  
GCCGTCGTTGGAAGTCTTTGATTAAGAAATATCCAAAACCTGAATCGGAGTTTGTCTATGGTGATTATAAAGT

FIG. 5 - CMV-T7-Cas9 D10A/H840A-3XFLAG-VP64:

TTATGATGTTCTGATAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACT  
CTAATATCATGAACCTTCTTCAAAACAGAAATTACACTTGCAAATGGAGAGATTTCGCAAACGCCCTCTAATCGAA  
ACTAATGGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCAGCAAAGTATTGTCC  
ATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAA  
AAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAATATGGTGGTTTTGATAGTCC  
AACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAATCCGTTAA  
AGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAATCCGATTGACTTTTTAGAAAGCTAAA  
GGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGTC  
GTAAACGGATGCTGGCTAGTGCCGGAGAATTACAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGA  
ATTTTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTT  
GTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTAGC  
AGATGCCAATTTAGATAAAGTTCCTAGTGATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAA  
AATATTATTCAATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGAT  
CGTAAACGATATACGCTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGA  
AACACGCATTGATTTGAGTCAGCTAGGAGGTGACGGTCTCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTA  
CAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGGctgcaggaggcgg  
tggaagcGGGCGCGCCGACGCGCTGGACGATTCGATCTCGACATGCTGGGTTCTGATGCCCTCGATGACTTTG  
ACCTGGATATGTTGGGAAGCGACGCATTGGATGACTTTGATCTGGACATGCTCGGCTCCGATGCTCTGGACGA  
TTTCGATCTCGATATGTTATAAccggtCATCATCACCATCACCATTGAGTTTAAACCCGCTGATCAGCCTCGACTG  
TGCCTTCTAGTTGCCAGCCATCTGTTGTTGCCCTCCCCCGTGCTTCTTGACCCTGGAAGGTGCCACTCCCA  
CTGTCTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGG  
GGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCT  
ATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCGATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCA  
TGGTCATAGCTGTTTCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAA  
GTGTAAGCCCTAGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTCCAGT  
CGGGAAACCTGTCGTGCCAGCTGCATTAAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCATATTGGGC  
GCTCTTCCGCTTCTCGCTCACTGACTCGCTGCGCTCGGTGCTTCCGCTGCGGCGAGCGGTATCAGCTCACTCA  
AAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCA  
AAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCAC  
AAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGA  
AGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTCTCCCTCGGGAAGC  
GTGGCGCTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGCTTCCGCTCCAAGCTGGGCTGTGT  
GCACGAACCCCCGTTAGCCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGA  
CACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACA  
GAGTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGC  
CAGTTACCTTCGGAAGAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTT  
TGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCT  
GACGCTCAGTGGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGA  
TCCTTTTAAATAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAACTTGGTCTGACAGTTACCAAT  
GCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGT  
AGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACC  
GGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCTCTGCAACTTTATC  
CGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAAC  
GTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCTGTTGGTATGGCTTCATTAGCTCCGGTCCCA  
ACGATCAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCTCCGATCGTT  
GTCAGAAGTAAGTTGGCCGAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTTACTGTATGCC  
ATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCG  
AGTTGCTCTTGCCGCGCTCAATACGGGATAATACCGCGCCACATAGCAGAAGTTTAAAGTGCTCATCATTG  
GAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCG  
TGCACCCAACCTGATCTTCAGCATCTTTTACTTTACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAAT

**FIG. 5 - CMV-T7-Cas9 D10A/H840A-3XFLAG-VP64:**

GCCGCAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTCAATATTATTGAA  
GCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGT  
TCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCGACGGATCGGGAGATCGATCTCCCGATCCCCTAGGG  
TCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGT  
CGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCT  
GCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACT  
AGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGTTACATAACTTAC  
GGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCGCCATTGACGTCAATAATGACGTATGTTCCCAT  
GTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTAC  
ATCAAGTGTATCC (SEQ ID NO:108)

FIG. 6 - MV-T7-Cas9 recoded D10A/H840A-3XFLAG-VP64

ATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCAGTACATGAC  
 CTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGC  
 AGTACATCAATGGGCGTGGATAGCGGTTTGA CTACG GGGGATTTC CAAAGTCTCCACCCCAT TGACGTCAATGG  
 GAGTTTGT TTTGGCACC AAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATG  
 GGCGGTAGGCGTG TACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGAG  
 ATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACCATGGATAAAAAAGTATTCTATTGGTTTAGc  
 CATCGGCACTAATTCCGTTGGATGGGCTGT CATAACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTG  
 TTGGGGAACACAGACCGTCATTGATTA AAAAGAATCTTATCGGTGCCCTCCTATTTCGATAGTGGCGAAACGG  
 CAGAGGCGACTCGCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACCGAATATGTTACTTAC  
 AAGAAATTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGTTTGGAAAGATCCTTCCTTGTC  
 GAAGAGGACAAGAAACATGAACGGCACCCCATCTTGGAAACATAGTAGATGAGGTGGCATATCATGAAAAG  
 TACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGGACCTGAGGTTAATCTACTT  
 GGCTCTTGCCCATATGATAAAGTTCGTTGGGCACTTCTCATTGAGGGTGATCTAAATCCGGACAACCTCGGAT  
 GTCGACAAACTGTTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTGAAGAGAACCCTATAAATGCAAGTG  
 GCGTGGATGCGAAGGCTATTCTTAGCGCCCGCTCTTAAATCCCGACGGCTAGAAAACCTGATCGCACAAAT  
 ACCCGGAGAGAAGAAAAATGGGTTGTTGCGTAACCTTAGCGCTCTCACTAGGCTGACACCAAATTTAAG  
 TCGAACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCGACAATC  
 TACTGGCACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCCTA  
 TCTGACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTACGATG  
 AACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATATT  
 CTTTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTT  
 TATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTA AAACTCAATCGCGAAGATCTACT  
 GCGAAAGCAGCGGACTTTCGACAACGGTAGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTT  
 AGAAGGCAGGAGGATTTTTATCCGTTCTCAAAGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTTCGCA  
 TACCTTACTATGTGGGACCCCTGGCCCGAGGGAACTCTCGGTTTCGCATGGATGACAAGAAAGTCCGAAGAAA  
 CGATTACTCCATGGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGAC  
 CAACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTTCACAGTGT  
 ACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAACCCGCCTTTCTAAGCGGAGAACAGA  
 AGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTT  
 TAAGAAAATTGAATGCTTCGATTCTGTCGAGATCTCCGGGGTAGAAGATCGATTTAATGCGTCACTTGGTACG  
 TATCATGACCTCCTAAAGATAATTAAGATAAGGACTTCCTGGATAACGAAGAGAATGAAGATATCTTAGAAG  
 ATATAGTGTTGACTCTTACCCTCTTGAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCTCACCT  
 GTTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGATTGTCGCGGAAACT  
 TATCAACGGGATAAGAGACAAGCAAAGTGGTAAAATATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAAT  
 AGGAACTTTATGCAGCTGATCCATGATGACTCTTAACTTCAAAGAGGATATACAAAAGGCACAGGTTTCCG  
 GACAAGGGGACTCATTGCACGAACATATTGCGAATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCATACTCCA  
 GACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAAACATTGTAATCGAGAT  
 GGCACGCGAAAATCAAACGACTCAGAAGGGGCAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAGAG  
 GGTATTAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTGCAGAACGAGAAA  
 CTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAGGAACTGGACATAAACCGTTTATCTGA  
 TTACGACGTCGATgcATTGTACCCCAATCCTTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTC  
 GGATAAGAACCGAGGGAAAAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGC  
 GGCAGCTCCTAAATGCGAACTGATAACGCAAAGAAAGTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCT  
 TGTCTGAAC TTGACAAGGCCGGATTTATTAACGTCA GCTCGTGGA AACCCGCCAAATCACAAGCATGTTGC  
 ACAGATACTAGATTC CGAATGAATACGAAATACGACGAGAACGATAAGCTGATT CGGGAAGTCAAAGTAAT  
 CACTTTAAAGTCAAAATTGGTGTGCGACTTCAGAAAGGATTTTCAATTCTATAAAGTTAGGGAGATAAATAACT  
 ACCACCATGCGCACGACGCTTATCTTAATGCCGTGCTAGGGACCGCACTCATTAGAAATACCCGAAGCTAGA  
 AAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCCGTAAGATGATCGCGAAAAGCGAACAGGAGAT

FIG. 6 - MV-T7-Cas9 recoded D10A/H840A-3XFLAG-VP64

AGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAATCACTCTGGCAA  
ACGGAGAGATACGCAAACGACCTTTAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGC  
CGGGACTTCGCGACGGTGAGAAAAGTTTTGTCCATGCCCAAGTCAACATAGTAAAGAAAAGTCTGAGGTGCAG  
ACCGGAGGGTTTTCAAAGGAATCGATTCTTCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGACT  
GGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCATTCTGTCTAGTAGTGGCAAAGTTGA  
GAAGGGAAAATCCAAGAACTGAAGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTT  
TGAAAAGAACCCCATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCATAATTAACTA  
CCAAAGTATAGTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGAGAGCTTCAAAAAG  
GGGAACGAACTCGCACTACCGTCTAAATACGTGAATTCCTGTATTTAGCGTCCCATTACGAGAAGTTGAAAG  
GTTACCTGAAGATAACGAACAGAACCACTTTTTGTTGAGCAGCACAAACATTATCTCGACGAAATCATAGA  
GCAAATTCGGAATTCAGTAAGAGAGTCATCCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAAC  
AAGCACAGGGATAAACCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCG  
CTCCAGCCGCAATCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGGAGGTGCTAGA  
CGCGACACTGATTCACCAATCCATCACGGGATTATATGAAACTCGGATAGATTTGTACAGCTTGGGGGTGAC  
GGATCCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTACAAAGACCATGACGGTGATTATAAAGATCATGAC  
ATCGATTACAAGGATGACGATGACAAGGctgcaggaggcgggtggaagcGGGCGCGCCGACGCGCTGGACGATTTG  
ATCTCGACATGCTGGGTTCTGATGCCCTCGATGACTTTGACCTGGATATGTTGGGAAGCGACGCATTGGATGA  
CTTTGATCTGGACATGCTCGGCTCCGATGCTCTGGACGATTTGATCTCGATATGTTATAAccggtCATCATCACC  
ATCACCATTGAGTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCT  
CCCCCGTGCTTCTTGACCTGGAAGGTGCCACTCCCACTGTCCTTCTCTAATAAAATGAGGAAATTGCATCG  
CATTGTCTGAGTAGGTGTCAATCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGA  
AGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCT  
CGATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCTGTGTGAAATTGTTATCCG  
CTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTAGGGTGCCTAATGAGTGAGCTAA  
CTCACATTAATTGCGTTGCGCTCACTGCCCCGTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAAT  
CGGCCAACGCGCGGGGAGAGGCGGTTTGCCTATTGGGCGCTCTTCCGCTTCTCGCTCACTGACTCGCTGCGC  
TCGGTCGTTGCGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGG  
GATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCT  
GGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCAAAAAATCGACGCTCAAGTCAGAGGTGGCGAAAC  
CCGACAGGACTATAAAGATACCAGGCGTTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCTGCC  
GCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGCGCTTTCTCAATGCTCACGCTGTAGGTATC  
TCAGTTCGGTGTAGGTGTTGCTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTAGCCCCAGCGCTGCGC  
CTTATCCGGTAACATATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGT  
AACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTAC  
ACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGAAAAAGAGTTGGTAGCTCTT  
GATCCGGCAAACAAACACCGCTGGTAGCGGTGGTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAA  
AAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAAGTACAGTTAAGG  
GATTTTGGTCATGAGATTATCAAAAAGGATCTTACCTAGATCCTTTTAAATTAATAAAGTATTAATCAA  
TCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATC  
TGTCTATTTGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATC  
TGCCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACGGCTCCAGATTTATCAGCAATAAACCAGCCA  
GCCGGAAGGGCCGAGCGCAGAAGTGGTCTGCAACTTTATCCGCTCCATCCAGTCTATTAATTGTTGCCGGG  
AAGCTAGAGTAAGTAGTTGCGCAGTTAATAGTTTGCACGTTGTTGCCATTGCTACAGGCATCGTGGTGTG  
ACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCTAACGATCAAGGCGAGTTACATGATCCCCATGT  
TGTGCAAAAAAGCGGTTAGCTCCTTCGGTCTCCGATCGTTGTGCAAGTAAGTTGGCCGAGTGTTATCACT  
CATGGTTATGGCAGCACTGCATAATTCTTACTGTATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGT  
ACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAA  
TACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAACGTTCTTCGGGGCGAAAAGTCTCAAGG  
ATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGACCCCACTGATCTTCAGCATCTTTTACTTTC  
ACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAGGGAATAAGGGCGACACGGAA

FIG. 6 - MV-T7-Cas9 recoded D10A/H840A-3XFLAG-VP64

ATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATA  
CATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCGAAAAAGTGCCACCTGAC  
GTCGACGGATCGGGAGATCGATCTCCCGATCCCCTAGGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCAT  
AGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTAC  
AACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGAT  
GTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGT  
TCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGAC  
CCCGGCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATG  
GGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATC (SEQ ID NO:109)

FIG. 7 - Cas9-activator protein

MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT  
 RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVD  
 EVAYHEKYPTIYHLRKKLVDSSTKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFI  
 QLVQTYNQLFEEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGL  
 TPNFKSNFDLAEDAKLQLSKDQYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT  
 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPPEKYKEIFFDQSKNGYAGYIDGGASQEEF  
 YKFIKPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPPFLK  
 DNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMT  
 NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK  
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIKDKDFLDNEENEDILEDIV  
 LTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF  
 LKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV  
 DELVKVMGRHKPENIVIEAMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL  
 QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVAIVPQSFLKDDSIDNKVLTRSDKNRGKSD  
 NVPSEEVVKKMKNYWRQLLNAKLITQRFKDNLTKAERGGLSELDKAGFIKRQLVETRQITKH  
 VAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVVREINNYHHAHDAYLNAV  
 VGTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEI GKATAYFFYSNIMNFFKTEITLAN  
 GEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNS  
 DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELLGITIMERSSSFENP  
 IDFLEAKGYKEVKKDLIIKLPKYSLFELNGRKRMLASAGELQKGNELALPSKYVNFLYLAS  
 HYEKLKGSPEQNEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKHDKPI  
 REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQ  
 LGGD**GS**pkkkrvssDYKDHDGDYKDHDIDYKDDDDKAAGGGGS**GRADALDDFDLMLGSDA**

**LDDFDLMLGSDALDDFDLMLGSDALDDFDLML** (SEQ ID NO:110)

FIG. 8A - dCas9-NLS-3XFLAG-HP1alpha

MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT  
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVD  
EVAYHEKYPTIYHLRKKLVDSTDKADLRLLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFI  
QLVQTYNQLFEEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNLFGNLIALSLGL  
TPNFKSNFDLAEDAKLQLSKDQYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT  
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF  
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLK  
DNREKIEKILTFRIPIYYVGPLARGNSRFAMWTRKSEETITPWNFEVVDKGASQSFIERMT  
NFDKNLPNEKVLPHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK  
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIV  
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWRLSRKLINGIRDKQSGKTILDF  
LKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVV  
DELVKVMGRHKPENIVIMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL  
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSD  
NVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH  
VAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAV  
VGTALIKKYPKLESEFVYGDKVYDVRKMIKSEQEI GKATAKYFFYSNIMNFFKTEITLAN  
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNVKKTETVQTGGFSKESILPKRNS  
DKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSSFENP  
IDFLEAKGYKEVKKDLIIKLPKYSLEFLENKRKMLASAGELQKGNELALPSKYVNFLYLAS  
HYEKLKGSPEQNEQQLFVEQHKHYLDEIIIEQISEFSKRVILADANLDKVL SAYNKH RDKPI  
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDL SQ  
LGGDGS[PKKKRKV]SSDYKDHDGDYKDHDIDYKDDDDKAAGGGGSMKEGENNKPREKSESNKR  
KSNFNSADDIKSKKKREQSNDIARGFERGLEPEKIIIGATDSCGDLMLMKWKDTDEADLVL  
AKEANVKCPQIVIAFYEEERLTWHAYPEDAENKEKETAKS (SEQ ID NO:111)

**FIG. 8B- dCas9-NLS-3XFLAG-HP1beta**

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT  
 RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVD  
 EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFI  
 QLVQTYNQLFEEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGL  
 TPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT  
 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF  
 YKFIKPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLLK  
 DNREKIEKILTFRIPIYYVGPLARGNSRFAMWTRKSEETITPWNFEEVVDKGASAQSFIERMT  
 NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK  
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIV  
 LTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF  
 LKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVV  
 DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL  
 QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSD  
 NVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH  
 VAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAV  
 VGTALIKKYPKLESEFVYGDKYVDVRKMIKSEQEI GKATAYFFYSNIMNFFKTEITLAN  
 GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIIVKKTEVQTGGFSKESILPKRNS  
 DKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSSFENP  
 IDFLEAKGYKEVKKDLIIKLPKYSLFELNGRKRMLASAGELQKGNELALPSKYVNFYLYLAS  
 HYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFSKRVILADANLDKVL SAYNKHDKPI  
 REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDL SQ  
 LGGDGSPKKKRKVSSDYKDHDGDYKDHDIDYKDDDDKAAGGGGSTTAHETDKSEGGKRKADSD  
SEDKGEESKPKKKKEESEKPRGFARGLEPERIIGATDSSGELMFLMKWKNSDDEADLVPAKEA  
NVKCPQVVISFYEERLTWHSYPSEDDDKKDDKN. (SEQ ID NO:112)

box = nuclear localization signal

underline = triple flag tag

double underline = HP1alpha hinge and chromoshadow domains

## FIG. 9- dCas9-3XFLAG-TET1CD

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT  
 RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLLEESFLVEEDKKHERHPIFGNIVD  
 EVAYHEKYPTIYHLRKKLVDSTDKADLRLLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFI  
 QLVQTYNQLFEEPNINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIALSLGL  
 TPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT  
 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF  
 YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLK  
 DNREKIEKILTFRIPYYVGPLARGNSRFAMWTRKSEETITPWNFEEVVDKGASAQSFIERMT  
 NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK  
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIV  
 LTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDQSGKTILDF  
 LKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV  
 DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL  
 QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVAIVPQSFLKDDSIDNKVLTSDKNRGKSD  
 NVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH  
 VAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAV  
 VGTALIKKYPKLESEFVYGDKYVDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLAN  
 GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIIVKKTEVQTGGFSKESILPKRNS  
 DKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSSFENP  
 IDFLEAKGYKEVKKDLIIKLPKYSLFELNGRKRMLASAGELQKGNELALPSKYVNFYLLAS  
 HYEKLGSPEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVIADANLDKVL SAYNKHDKPI  
 REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSIITGLYETRIDL  
 LSGDGSPKKKRKVSSDYKDHDGDYKDHDIDYKDDDDKAAGGGGSLPTCSCLD RVIQKDKGPY  
YTHLGAGPSVAAVREIMENRYGQKGNAIRIEIVVYTGKEGKSSHGCPIAKWVLRRSSDEEKV  
LCLVRQRTGHHCPTAVMVVLIMVWDGIPLPMADRLYTELTENLKSYNGHPTDRCTLNENRT  
CTCQGIDPETCGASFSFGCSWSMYFNGCKFGRSPSPRRFRIDPSSPLHEKNLEDNLQSLATR  
LAPIYKQYAPVAYQNOVEYENVARECRLGSKEGRPFSGVTACLDFCAHPHRDIHNMNNGSTV  
VCTLTTREDNRSLGVIPQDEQLHVLPLYKLSDTDEFGSKEGMEAKIKSGAIEVLAPRRKKRTC  
FTQPVPRSGKKRAAMMTEVLAHKIRAVEKKPIPRIKRKNNSTTTNNSKPSSLPTLGSNTETV  
QPEVKSETEPHILKSSDNTKTYSLMPSAPHPVKEASPGFSWSPKTASATPAPLKNDATAS  
GFSERSSTPHCTMPSGRLSGANAAADGPGISQLGEVAPLPTLSAPVMEPLINSEPSTGVTE  
PLTPHQPNHQPSFLTSPODLASSPMEEDEQHSEADEPPSDEPLSDDPLSPAEEKLPHIDEYW  
SDSEHIFLDANIGGVAIAPAHGSVLIECARRELHATTPVEHPNRNHPTRLSLVFYQHKNLNK  
POHGFELNKIKFEAKEAKNKMKASEQKDQAANEGPEQSSEVNELNQIPSHKALTLTHDNV  
TVSPYALTHVAGPYNHWV

(SEQ ID NO:113)

box = nuclear localization signal

underline = triple flag tag

double underline = TET1CD

19 / 20

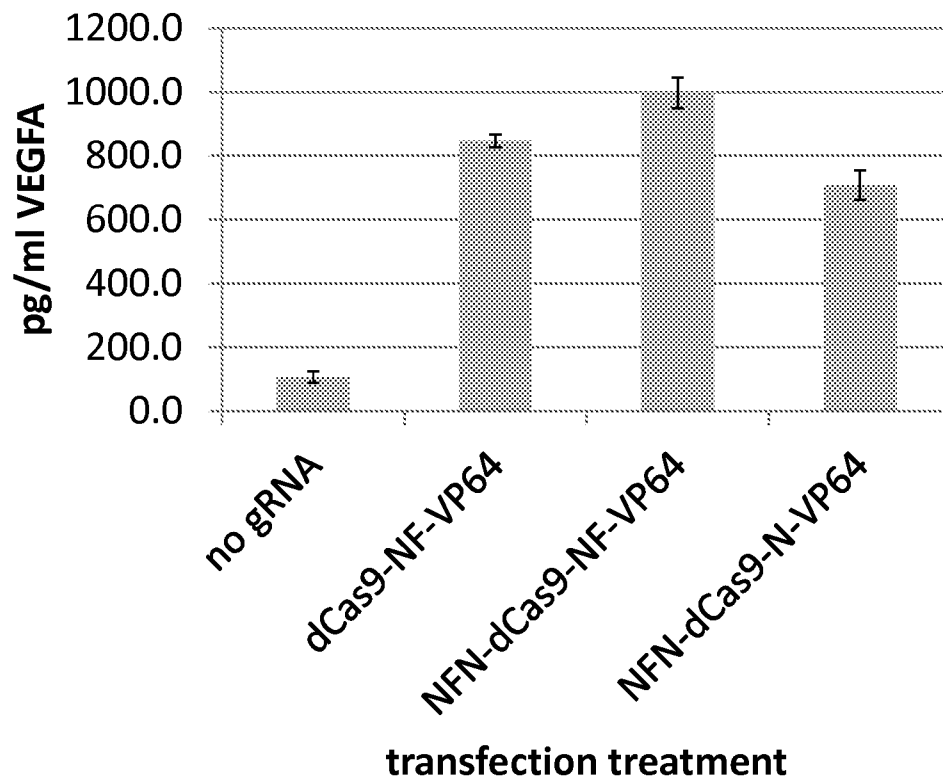


FIG. 10

20 / 20

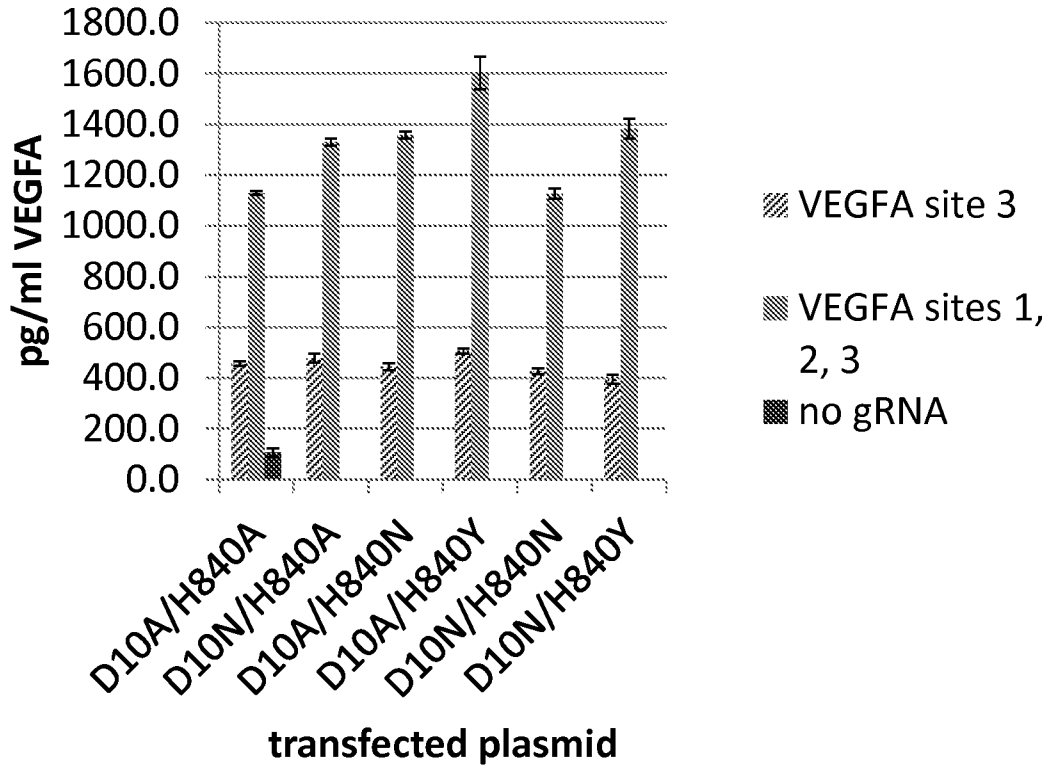


FIG. 11A

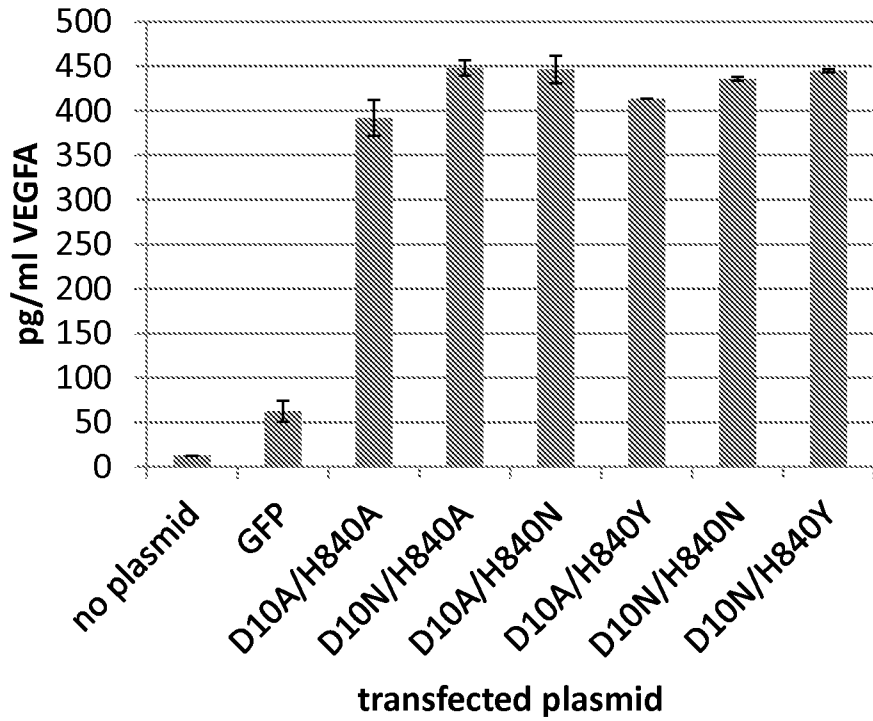


FIG. 11B