The binding of a ligand to the intein results in self-excision of the intein, restoring the activity of the protein.

Abstract: Some aspects of this disclosure provide compositions, methods, systems, and kits for controlling the activity of RNA-programmable endonucleases, such as Cas9, or for controlling the activity of proteins comprising a Cas9 variant fused to a functional effector domain, such as a nuclease, nickase, recombinase, deaminase, transcriptional activator, transcriptional repressor, or epigenetic modifying domain. For example, the inventive proteins provide comprise a ligand-dependent intein, the presence of which inhibits one or more activities of the protein (e.g., gRNA binding, enzymatic activity, target DNA binding). The binding of a ligand to the intein results in self-excision of the intein, restoring the activity of the protein.
CAS9 PROTEINS INCLUDING LIGAND-DEPENDENT INTEINS

RELATED APPLICATIONS


GOVERNMENT SUPPORT

[0002] This invention was made with U.S. Government support under R01 GM095501 and F32GM 106601, awarded by the National Institutes of Health/National Institute of General Medical Sciences, and under grant numbers HR001 1-1-2-0003 and N66001-12-C-4207, awarded by the Department of Defense. The Government has certain rights in the invention.

BACKGROUND OF THE INVENTION

[0003] Site-specific enzymes theoretically allow for the targeted manipulation of a single site within a genome and are useful in the context of gene targeting as well as for therapeutic applications. In a variety of organisms, including mammals, site-specific enzymes such as endonucleases have been used for genome engineering by stimulating either non-homologous end-joining or homologous recombination. In addition to providing powerful research tools, site-specific nucleases also have potential as gene therapy agents, and two site-specific endonucleases have recently entered clinical trials: one, CCR5-2246, targeting a human CCR-5 allele as part of an anti-HIV therapeutic approach (clinical trials NCT00842634, NCT01044654, NCT01252641), and the other one, VF24684, targeting the human VEGF-A promoter as part of an anti-cancer therapeutic approach (clinical trial NCT01082926).

[0004] Specific manipulation of the intended target site without or with only minimal off-target activity is a prerequisite for clinical applications of site-specific enzymes, and also for high-efficiency genomic manipulations in basic research applications. For example, imperfect specificity of engineered site-specific binding domains of certain nucleases has been linked to cellular toxicity and undesired alterations of genomic loci other than the intended target. Most nucleases available today, however, exhibit significant off-target
activity, and thus may not be suitable for clinical applications. An emerging nuclease platform for use in clinical and research settings are the RNA-guided nucleases, such as Cas9. While these nucleases are able to bind guide RNAs (gRNAs) that direct cleavage of specific target sites, off-target activity is still observed for certain Cas9:gRNA complexes (Pattanayak et al., "High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity." Nat Biotechnol. 2013; doi: 10.1038/nbt.2673). Technology for engineering site-specific enzymes with reduced off-target effects is therefore needed.

**SUMMARY OF THE INVENTION**

[0005] The reported toxicity of some engineered site-specific enzymes such as endonucleases is thought to be based on off-target DNA cleavage. Further, the activity of existing RNA-guided nucleases generally cannot be controlled at the molecular level, for example, to switch a nuclease from an "off" to an "on" state. Controlling the activity of nucleases and other site-specific enzymes suitable for nucleic acid manipulations or modifications could decrease the likelihood of incurring off-target effects. Some aspects of this disclosure provide strategies, compositions, systems, and methods to control the binding and/or enzymatic activity of RNA-programmable enzymes, such as Cas9 endonuclease, nickases, deaminases, recombinases, transcriptional activators and repressors, epigenetic modifiers variants and fusions thereof.

[0006] Accordingly, one aspect of the present disclosure provides Cas9 proteins (including fusions of Cas9 proteins and functional domains) comprising inteins, for example, ligand-dependent inteins. The presence of the intein inhibits one or more activities of the Cas9 proteins, for example, nucleic acid binding activity (e.g., target nucleic acid binding activity and/or gRNA binding activity), a nuclease activity, or another enzymatic activity (e.g., nucleic acid modifying activity, transcriptional activation and repression, etc.) for which the Cas9 protein (e.g., Cas9 fusion protein) is engineered to undertake (e.g., nuclease activity, nickase activity, recombinase activity, deaminase activity, transcriptional activator/repressor activity, epigenetic modification, etc.). In some embodiments, the Cas9 protein is a Cas9 nickase. The Cas9 fusions are typically between a nuclease inactivated Cas9 ("dCas") and one or more functional domains. The intein may be inserted into any location of a Cas9 protein, including one or more domains of a Cas9 protein or Cas9 fusion (including in a functional domain), such as the HNH nuclease domain or the RuvC nuclease domain. In some embodiments, the intein replaces amino acid residue Cys80, Ala127, Thr146, Ser219, Thr333, Thr519, Cys574, Thr622, Ser701, Ala728, Thr995, Ser1006,
Serl 154, Serl 159, or Serl274 in the Cas9 polypeptide sequence set forth as SEQ ID NO:2, in the dCas9 polypeptide sequence set forth as SEQ ID NO:5, or in the Cas9 nickase polypeptide sequence set forth as SEQ ID NO:4. In some embodiments, the intein replaces or is inserted at an amino acid residue that is within 5, within 10, within 15, or within 20 amino acid residues of Cys80, Alal27, Thrl46, Ser219, Thr333, Thr519, Cys574, Thr622, Ser701, Ala728, Thr995, Serl006, Serl 154, Serl 159, or Serl274 in the Cas9 polypeptide sequence set forth as SEQ ID NO:2, in the dCas9 polypeptide sequence set forth as SEQ ID NO:5, or in the Cas9 nickase polypeptide sequence set forth as SEQ ID NO:4. The intein replaces amino acid residue Alal27, Thrl46, Ser219, Thr519, or Cys574 in the Cas9 polypeptide sequence set forth as SEQ ID NO:2, in the dCas9 polypeptide sequence set forth as SEQ ID NO:5, or in the Cas9 nickase polypeptide sequence set forth as SEQ ID NO:4. Typically the intein is a ligand-dependent intein which exhibits no or minimal protein splicing activity in the absence of ligand (e.g., small molecules such as 4-hydroxytamoxifen, peptides, proteins, polynucleotides, amino acids, and nucleotides). Ligand-dependent inteins are known, and include those described in U.S. patent application, U.S.S.N. 14/004,280, published as U.S. 2014/0065711 Al, the entire contents of which are incorporated herein by reference. In some embodiments, the intein comprises an amino acid sequence selected from the group consisting of SEQ ID NO:7-14.

[0007] In one aspect, a Cas9 protein is provided that comprises: (i) a nuclease-inactivated Cas9 (e.g., dCas9 (SEQ ID NO:5)) domain; (ii) a ligand-dependent intein; and (iii) a recombinase catalytic domain. In some embodiments, the ligand-dependent intein domain is inserted into the dCas9 domain as described herein. Typically, the presence of the intein in the Cas9 protein inhibits one or more activities of the Cas9 protein, such as gRNA binding activity, target nucleic acid binding activity, and/or recombinase activity. Accordingly, upon self-excision of the intein (e.g., induced by ligand binding the intein) the one or more activities of the Cas9 protein is/are restored. In some embodiments, the recombinase catalytic domain is a monomer of the recombinase catalytic domain of Hin recombinase, Gin recombinase, or Tn3 recombinase.

[0008] According to another aspect, a Cas9 protein is provided that comprises: (i) a nuclease-inactivated Cas9 (e.g., dCas9 (SEQ ID NO:5)) domain; (ii) a ligand-dependent intein; and (iii) a deaminase catalytic domain. In some embodiments, the ligand-dependent intein domain is inserted into the dCas9 domain as described herein. Typically, the presence of the intein in the Cas9 protein inhibits one or more activities of the Cas9 protein, such as gRNA binding activity, target nucleic acid binding activity, and/or deaminase activity.
Accordingly, upon self-excision of the intein (e.g., induced by ligand binding of the intein) the one or more activities of the Cas9 protein is/are restored. In some embodiments, the deaminase catalytic domain comprises a cytidine deaminase (e.g., of apolipoprotein B niRNA-editing complex (APOBEC) family deaminases such as APOBEC1 or activation-induced cytidine deaminase (AID)). In some embodiments, the deaminase catalytic domain comprises a ACF1/ASE deaminase or an adenosine deaminase, such as a ADAT family deaminase.

According to another aspect, a Cas9 protein is provided that comprises: (i) a nuclease-inactivated Cas9 (e.g., dCas9 (SEQ ID NO:5)) domain; (ii) a ligand-dependent intein; and (iii) a transcriptional activator domain. In some embodiments, the ligand-dependent intein domain is inserted into the dCas9 domain as described herein. Typically, the presence of the intein in the Cas9 protein inhibits one or more activities of the Cas9 protein, such as gRNA binding activity, target nucleic acid binding activity, and/or transcriptional activation. Accordingly, upon self-excision of the intein (e.g., induced by ligand binding the intein) the one or more activities of the Cas9 protein is/are restored. In some embodiments, the transcriptional activator domain is VP64, CP16, and p65.

According to another aspect, a Cas9 protein is provided that comprises: (i) a nuclease-inactivated Cas9 (e.g., dCas9 (SEQ ID NO:5)) domain; (ii) a ligand-dependent intein; and (iii) a transcriptional repressor domain. In some embodiments, the ligand-dependent intein domain is inserted into the dCas9 domain as described herein. Typically, the presence of the intein in the Cas9 protein inhibits one or more activities of the Cas9 protein, such as gRNA binding activity, target nucleic acid binding activity, and/or transcriptional repression. Accordingly, upon self-excision of the intein (e.g., induced by ligand binding the intein) the one or more activities of the Cas9 protein is/are restored. In some embodiments, the transcriptional repressor domain is KRAB, SID, or SID4x.

According to another aspect, a Cas9 protein is provided that comprises: (i) a nuclease-inactivated Cas9 (e.g., dCas9 (SEQ ID NO:5)) domain; (ii) a ligand-dependent intein; and (iii) an epigenetic modifier domain. In some embodiments, the ligand-dependent intein domain is inserted into the dCas9 domain as described herein. Typically, the presence of the intein in the Cas9 protein inhibits one or more activities of the Cas9 protein, such as gRNA binding activity, target nucleic acid binding activity, and/or epigenetic modification activity. Accordingly, upon self-excision of the intein (e.g., induced by ligand binding the intein) the one or more activities of the Cas9 protein is/are restored. In some embodiments, the epigenetic modifier domain is epigenetic modifier is selected from the group consisting of...
histone demethylase, histone methyltransferase, hydroxylase, histone deacetylase, and histone acetyltransferase. In some embodiments, the epigenetic modifier comprises the LSD1 histone demethylase or TET1 hydroxylase.

According to another aspect, methods of using Cas9 proteins are provided. In some embodiments involving site-specific DNA cleavage, the methods comprise (a) contacting a Cas9 protein (e.g., having nuclease activity) comprising a ligand-dependent intein with a ligand, wherein binding of the ligand to the intein induces self-excision of the intein; and (b) contacting a DNA with the Cas9 protein, wherein the Cas9 protein is associated with a gRNA; whereby self-excision of the intein from the Cas9 protein in step (a) allows the Cas9 protein to cleave the DNA, thereby producing cleaved DNA. In some embodiments, the Cas9 protein first binds a gRNA and optionally the target DNA prior to excision of the intein, but is unable to cleave the DNA until excision of the intein occurs. Any of the Cas9 proteins having nuclease activity and comprising a ligand-dependent intein, as described herein, can be used in the inventive methods.

According to another aspect, methods of using any of the ligand-dependent intein-containing Cas9 proteins comprising a recombinase catalytic domain are provided. In some embodiments, the method is useful for recombining two nucleic acids, such as two DNAs, and comprises (a) contacting a first DNA with a first ligand-dependent dCas9-recombinase fusion protein (e.g., any of those described herein), wherein the dCas9 domain of the first fusion protein binds a first gRNA that hybridizes to a region of the first DNA; (b) contacting the first DNA with a second ligand-dependent dCas9-recombinase fusion protein, wherein the dCas9 domain of the second fusion protein binds a second gRNA that hybridizes to a second region of the first DNA; (c) contacting a second DNA with a third ligand-dependent dCas9-recombinase fusion protein, wherein the dCas9 domain of the third fusion protein binds a third gRNA that hybridizes to a region of the second DNA; and (d) contacting the second DNA with a fourth ligand-dependent dCas9-recombinase fusion protein, wherein the dCas9 domain of the fourth fusion protein binds a fourth gRNA that hybridizes to a second region of the second DNA; whereby the binding of the fusion proteins in steps (a) - (d) results in the tetramerization of the recombinase catalytic domains of the fusion proteins, under conditions such that the DNAs are recombined. In some embodiments, the methods are useful for site-specific recombination between two regions of a single DNA molecule, and comprise (a) contacting the DNA with a first ligand-dependent dCas9-recombinase fusion protein, wherein the dCas9 domain if the first fusion protein binds a first gRNA that hybridizes to a region of the DNA; (b) contacting the DNA with a second ligand-dependent
dCas9-recombinase fusion protein, wherein the dCas9 domain of the second fusion protein binds a second gRNA that hybridizes to a second region of the DNA; (c) contacting the DNA with a third ligand-dependent dCas9-recombinase fusion protein, wherein the dCas9 domain of the third fusion protein binds a third gRNA that hybridizes to a third region of the DNA; (d) contacting the DNA with a fourth ligand-dependent dCas9-recombinase fusion protein, wherein the dCas9 domain of the fourth fusion protein binds a fourth gRNA that hybridizes to a fourth region of the DNA; whereby the binding of the fusion proteins in steps (a) - (d) results in the tetramerization of the recombinase catalytic domains of the fusion proteins, under conditions such that the DNA is recombined. In some embodiments, any of the methods first comprise contacting the fusion proteins with a ligand that induces self-excision of the intein. In some embodiments, the fusion proteins are contacted with the ligand after: (i) the fusion proteins bind a gRNA; (ii) the fusion proteins bind the DNA; or (iii) after the recombinase domains form a tetramer. In some embodiments, the gRNAs in any step (a)-(d) of the inventive methods hybridize to the same strand or to opposing strands in the DNA(s). In some embodiments, the gRNAs hybridize to regions of their respective DNAs that are no more than 10, no more than 15, no more than 20, no more than 25, no more than 30, no more than 40, no more than 50, no more than 60, no more than 70, no more than 80, no more than 90, or no more than 100 base pairs apart.

[0013] According to yet another aspect, methods of using any of the ligand-dependent intein Cas9 proteins comprising deaminase catalytic domains are provided. The methods comprise contacting a DNA molecule with (a) a ligand-dependent Cas9 protein comprising deaminase catalytic domain as provided herein; and (b) a gRNA targeting the Cas9 protein of step (a) to a target nucleotide sequence of the DNA strand; wherein the DNA molecule is contacted with the Cas9 protein, and the gRNA in an amount effective and under conditions suitable for the deamination of a nucleotide base. In some embodiments, the methods comprise contacting the Cas9 protein with a ligand that induces self-excision of the intein either before or after the Cas9 protein binds the gRNA. In some embodiments, the target DNA sequence comprises a sequence associated with a disease or disorder, and wherein the deamination of the nucleotide base results in a sequence that is not associated with a disease or disorder. In some embodiments, the DNA sequence to be modified comprises a T→C or A→G point mutation associated with a disease or disorder, and the deamination of the mutant C or G base results in a sequence that is not associated with a disease or disorder (e.g., the deamination corrects the mutation the caused the disease or disorder). In some embodiments, the deamination corrects a point mutation in the sequence associated with the disease or
disorder. In some embodiments, the sequence associated with the disease or disorder encodes a protein, and wherein the deamination introduces a stop codon into the sequence associated with the disease or disorder, resulting in a truncation of the encoded protein. In some embodiments, the deamination corrects a point mutation in the PI3KCA gene, thus correcting an H1047R and/or a A3140G mutation. In some embodiments, the contacting is performed in vivo in a subject susceptible to having or diagnosed with the disease or disorder. In some embodiments, the disease or disorder is a disease associated with a point mutation, or a single-base mutation, in the genome. In some embodiments, the disease is a genetic disease, a cancer, a metabolic disease, or a lysosomal storage disease.

[0014] According to another aspect, methods for transcriptional activation of a gene are provided. In some embodiments, the methods comprise contacting a DNA molecule comprising a gene with (a) a ligand-dependent dCas9 fusion protein comprising a transcriptional activator (e.g., any of those provided herein) and (b) a gRNA targeting the fusion protein of (a) to a target nucleotide sequence of the DNA strand; wherein the DNA molecule is contacted with the fusion protein and the gRNA in an amount effective and under conditions suitable for the transcriptional activation of the gene. In some embodiments, the method further comprises contacting the fusion protein with a ligand that induces self-excision of the intein. In some embodiments, the fusion protein is contacted with the ligand prior to forming a complex with a gRNA. In some embodiments, the fusion protein is contacted with the ligand after forming a complex with a gRNA. In some embodiments, the gRNA targets the promoter region of a gene.

[0015] According to another aspect, methods for transcriptional repression of a gene are provided. In some embodiments, the methods comprise contacting a DNA molecule comprising a gene with (a) a ligand-dependent dCas9 fusion protein comprising a transcriptional repressor (e.g., any of those provided herein) and (b) a gRNA targeting the fusion protein of (a) to a target nucleotide sequence of the DNA strand; wherein the DNA molecule is contacted with the fusion protein and the gRNA in an amount effective and under conditions suitable for the transcriptional repression of the gene. In some embodiments, the method further comprises contacting the fusion protein with a ligand that induces self-excision of the intein. In some embodiments, the fusion protein is contacted with the ligand prior to forming a complex with a gRNA. In some embodiments, the fusion protein is contacted with the ligand after forming a complex with a gRNA. In some embodiments, the gRNA targets the promoter region of a gene.
[0016] According to another aspect, methods for epigenetic modification of DNA are provided. In some embodiments, the DNA is chromosomal DNA. In some embodiments, the methods comprise contacting a DNA molecule with (a) a ligand-dependent dCas9 fusion protein comprising a epigenetic modifier (e.g., any of those provided herein) and (b) a gRNA targeting the fusion protein of (a) to a target nucleotide sequence of the DNA strand; wherein the DNA molecule is contacted with the fusion protein and the gRNA in an amount effective and under conditions suitable for the epigenetic modification of the DNA. In some embodiments, the method further comprises contacting the fusion protein with a ligand that induces self-excision of the intein. In some embodiments, the fusion protein is contacted with the ligand prior to forming a complex with a gRNA. In some embodiments, the fusion protein is contacted with the ligand after forming a complex with a gRNA. In some embodiments, the gRNA targets the promoter region of a gene in the DNA.

[0017] Any of the methods provided herein can be performed on DNA in a cell, for example, a cell in vitro or in vivo. In some embodiments, any of the methods provided herein are performed on DNA in a eukaryotic cell. In some embodiments, the eukaryotic cell is in an individual, for example, a human.

[0018] According to some embodiments, polynucleotides are provided, for example, that encode any of the proteins (e.g., proteins comprising ligand-dependent Cas9 proteins or variants) described herein. In some embodiments, vectors that comprise a polynucleotide described herein are provided. In some embodiments, vectors for recombinant expression of any of the proteins (e.g., comprising ligand-dependent Cas9 proteins or variants) described herein are provided. In some embodiments, cells comprising genetic constructs for expressing any of the proteins (e.g., comprising ligand-dependent Cas9 proteins or variants) described herein are provided.

[0019] In some embodiments, kits useful in using, producing, or creating any of the ligand-dependent Cas9 proteins or variants thereof, as described herein, are provided. For example, kits comprising any of the proteins (e.g., ligand-dependent Cas9 proteins or variants) described herein are provided. In some embodiments, kits comprising any of the polynucleotides described herein are provided. In some embodiments, kits comprising a vector for recombinant expression, wherein the vectors comprise a polynucleotide encoding any of the proteins (e.g., ligand-dependent Cas9 proteins or variants) described herein, are provided. In some embodiments, kits comprising a cell comprising genetic constructs for expressing any of the proteins (e.g., ligand-dependent Cas9 proteins or variants) described herein are provided.
[0020] Other advantages, features, and uses of the invention will be apparent from the Detailed Description of Certain Embodiments of the Invention; the Drawings, which are schematic and not intended to be drawn to scale; and the Claims.

**BRIEF DESCRIPTION OF THE DRAWINGS**

[0021] Figure 1 shows a schematic depicting an exemplary embodiment of the disclosure. A Cas9 protein comprising a ligand-dependent intein, remains inactive in the absence of a ligand that binds the intein domain. Upon addition of the ligand, the intein is self-excised, restoring the activity of the Cas9 protein. Cas9 is then able to mediate RNA-guided cleavage of a DNA target sequence.

[0022] Figure 2 shows the results of T7 Endonuclease I Surveyor assay used to assess ligand-dependent Cas9 gene modification at three target sites (EMX, VEGF, or CLTA). The presence of two bands corresponding to smaller DNA fragments (the fragments are approximately the same size for EMX) indicates genomic modification.

[0023] Figure 3A-C. Insertion of an evolved ligand-dependent intein enables small-molecule control of Cas9. (A) Intein insertion renders Cas9 inactive. Upon 4-HT binding, the intein undergoes conformational changes that trigger protein splicing and restore Cas9 activity. (B) The evolved intein was inserted to replace each of the colored residues. Intein-inserted Cas9 variants at S219 and C574 (green) were used in subsequent experiments. (C) Genomic EGFP disruption activity of wild-type Cas9 and intein-Cas9 variants in the absence or presence of 4-HT. Intein-Cas9 variants are identified by the residue replaced by the intein. Error bars reflect the standard deviation of three biological replicates.

[0024] Figure 4A-D. Genomic DNA modification by intein-Cas9(S219), intein-Cas9(C574), and wild-type Cas9. (A) Indel frequency from high-throughput DNA sequencing of amplified genomic on-target sites in the absence or presence of 4-HT. Note that a significant number of indels were observed at the CLTA on-target site even in the absence of a targeting sgRNA (Table 9). (B-D) DNA modification specificity, defined as on-target:off-target indel frequency ratio$^{46}$, normalized to wild-type Cas9. Cells were transfected with 500 ng of the Cas9 expression plasmid. P-values are $<10^{-13}$ for the Fisher exact test (one-sided up) on comparisons of indel modification frequency in the presence versus the absence of 4-HT for intein-Cas9(S219) and intein-Cas9(C574). P-values were adjusted for multiple comparisons using the Benjamini-Hochberg method, and are listed in Table 5. Error bars reflect the range of two independent experiments conducted on different days.
Figure 5. Effect of 4-HT on cellular toxicity. Untransfected HEK293-GFP stable cells, and cells transfected with intein-Cas9(S219) and sgRNA expression plasmids, were treated with or without 4-HT (1 µM). 12 h after transfection, the media was replaced with full serum media, with or without 4-HT (1 µM). Cells were thus exposed to 4-HT for 0, 12, or 60 h. The live cell population was determined by flow cytometry 60 h after transfection using TO-PRO-3 stain (Life Technologies). Error bars reflect the standard deviation of six technical replicates.

Figure 6A-B. Western blot analysis of HEK293-GFP stable cells transfected with (A) wild-type Cas9 or (B) intein-Cas9(S219) expression plasmid. 12 h after transfection, cells were treated with or without 4-HT (1 µM). Cells were lysed and pooled from three technical replicates 4, 8, 12, or 24 h after 4-HT treatment. An anti-FLAG antibody (Sigma-Aldrich F1804) and an anti-mouse 800CW IRDye (LI-COR) were used to visualize the gel. Lanes 1 and 2 contain purified dCas9-VP64-3xFLAG protein and lysate from untransfected HEK293 cells, respectively.

Figure 7. Indel frequency from high-throughput DNA sequencing of amplified genomic on-target sites ("On") and off-target sites ("Off 1-Off 4") by intein-Cas9(S219), intein-Cas9(C574), and wild-type Cas9 in the presence of 4-HT. 500 ng of Cas9 expression plasmid was transfected. The higher observed efficiency of VEGF Off 1 modification than VEGF on-target modification is consistent with a previous report. P-values are < 0.005 for the Fisher exact test (one-sided down) on all pairwise comparisons within each independent experiment of off-target modification frequency between either intein-Cas9 variant in the presence of 4-HT versus that of wild-type Cas9 in the presence of 4-HT. P-values were adjusted for multiple comparisons using the Benjamini-Hochberg method, and are listed in Table 7. Error bars reflect the range of two independent experiments conducted on different days. See also Fu, Y. et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nature biotechnology 31, 822-826 (2013).

Figure 8A-C. DNA modification specificity of intein-Cas9(S219), intein-Cas9(C574), and wild-type Cas9 in the absence of 4-HT. (A-C) On-target:off-target indel frequency ratio following the transfection of 500 ng of intein-Cas9(S219), intein-Cas9(C574), or wild-type Cas9 expression plasmid.

Figure 9. Genomic on-target DNA modification by intein-Cas9(S219), intein-Cas9(C574), and wild-type Cas9 in the presence of 4-HT. Five different amounts of wild-type Cas9 expression plasmid, specified in parenthesis, were transfected. P-values for
comparisons between conditions (Table 8) were obtained using the Fisher exact test and
adjusted for multiple comparisons using the Benjamini-Hochberg Method.

[0030] Figure 10A-B. Indel frequency from high-throughput DNA sequencing of
amplified genomic on-target sites ("On") and off-target sites ("Off 1-Off 4") by intein-
Cas9(S219), intein-Cas9(C574), and wild-type Cas9 in the presence of 4-HT. Five different
amounts of wild-type Cas9 expression plasmid, specified in parenthesis, were transfected (A).
Genomic sites with low modification frequencies are enlarged in (B). P-values for
comparisons between conditions (Table 8) were obtained using the Fisher exact test and
adjusted for multiple comparisons using the Benjamini-Hochberg Method.

[0031] Figure 11A-C. DNA modification specificity of intein-Cas9(S219), intein-
Cas9(C574), and wild-type Cas9 in the presence of 4-HT. (A-C) On-target: off-target indel
frequency ratio normalized to wild-type Cas9 (500 ng). Five different amounts of wild-type
Cas9 expression plasmid, specified in parenthesis, were transfected.

[0032] Figure 12A-B. Genomic EGFP disruption activity of intein-Cas9(S219) and
intein-Cas9(S219-G521R) in the presence of (A) β-estradiol or (B) 4-HT. Error bars reflect
the standard deviation of three technical replicates.

DEFINITIONS

[0033] As used herein and in the claims, the singular forms "a," "an," and "the"
include the singular and the plural reference unless the context clearly indicates otherwise.
Thus, for example, a reference to "an agent" includes a single agent and a plurality of such
agents.

[0034] The term "Cas9" or "Cas9 nuclease" refers to an RNA-guided nuclease
comprising a Cas9 protein, or a fragment thereof. A Cas9 nuclease is also referred to
sometimes as a cas9 nuclease or a CRISPR (clustered regularly interspaced short
palindromic repeat)-associated nuclease. CRISPR is a prokaryotic adaptive immune system
that provides protection against mobile genetic elements (e.g., viruses, transposable elements,
and conjugative plasmids). CRISPR clusters contain spacers, sequences complementary to
antecedent mobile elements, and target invading nucleic acids. CRISPR clusters are
transcribed and processed into CRISPR RNA (crRNA). In type II CRISPR systems correct
processing of pre-crRNA requires a trans-encoded small RNA (tracrRNA), endogenous
ribonuclease 3 (rnc), and a Cas9 protein. The tracrRNA serves as a guide for ribonuclease 3-
aided processing of pre-crRNA. Subsequently, Cas9/crRNA/tracrRNA endonucleolytically
cleaves linear or circular dsDNA target complementary to the spacer. The target strand not
complementary to crRNA is first cut endonucleolytically, then trimmed 3' -> 5'
exonucleolytically. In nature, DNA-binding and cleavage typically requires protein and both RNA species. However, single guide RNAs ('sgRNA', or simply "gRNA") can be engineered so as to incorporate aspects of both the crRNA and tracrRNA into a single RNA molecule. See, e.g., Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. Science 337:816-821(2012), the entire contents of which is hereby incorporated by reference. Cas9 recognizes a short motif in the CRISPR repeat sequences (the PAM or protospacer adjacent motif) to help distinguish self versus non-self. Cas9 nuclease sequences and structures are well known to those of skill in the art (see, e.g., "Complete genome sequence of an M1 strain of Streptococcus pyogenes." Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L. expand/collapse author list McLaughlin R.E., Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma CM., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607(2011) and "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. Science 337:816-821(2012), the entire contents of each of which are incorporated herein by reference). Cas9 orthologs have been described in various species, including, but not limited to, S. pyogenes and S. thermophilus. Additional suitable Cas9 nucleases and sequences will be apparent to those of skill in the art based on this disclosure, and such Cas9 nucleases and sequences include Cas9 sequences from the organisms and loci disclosed in Chylinski, Rhun, and Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) RNA Biology 10:5, 726-737; the entire contents of which are incorporated herein by reference. In some embodiments, proteins comprising Cas9 proteins or fragments thereof are referred to as "Cas9 variants." A Cas9 variant shares homology to Cas9, or a fragment thereof. For example, a Cas9 variant may be at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to wild type Cas9. In some embodiments, the Cas9 variant comprises a fragment of Cas9 (e.g., a gRNA binding domain or a DNA-cleavage domain, an N-terminal domain or a C-terminal domain, etc.), such that the fragment is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5%
identical, or at least about 99.9% to the corresponding fragment of wild type Cas9. In some embodiments, wild type Cas9 corresponds to Cas9 from *Streptococcus pyogenes* (NCBI Reference Sequences: NC_017053.1 and NC_002737.1). In some embodiments, wild type Cas9 corresponds to SEQ ID NO:1 (nucleotide); SEQ ID NO:2 (amino acid)). In some embodiments, wild type Cas9 corresponds to a human codon optimized sequence of Cas9 (e.g., SEQ ID NO:3; See, e.g., Fu *et al*. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol*. 2013; **31**, 822-826). In some embodiments, a Cas9 nuclease has an inactive (e.g., an inactivated) DNA cleavage domain. A nuclease-inactivated Cas9 protein may also be referred to as a "dCas9" protein (for nuclease "dead" Cas9). In some embodiments, dCas9 corresponds to, or comprises in part or in whole, the amino acid set forth as SEQ ID NO:5, below. In some embodiments, variants of dCas9 (e.g., variants of SEQ ID NO:5) are provided. For example, in some embodiments, variants having mutations other than D10A and H840A are provided, which *e.g.*, result in a nuclease inactivated Cas9 (dCas9). Such mutations, by way of example, include other amino acid substitutions at D10 and H840, or other substitutions within the nuclease domains of Cas9 (e.g., substitutions in the HNH nuclease subdomain and/or the RuvCl subdomain). In some embodiments, a Cas9 protein variant is a Cas9 nickase, which includes a mutation which abolishes the nuclease activity of one of the two nuclease domains of the protein. In some embodiments, a Cas9 nickase has one, but not both of a D10A and H840A substitution. In some embodiments, a Cas9 nickase corresponds to, or comprises in part or in whole, the amino acid set forth as SEQ ID NO:4, below. In some embodiments, variants or homologues of dCas9 or Cas9 nickase are provided which are at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to SEQ ID NO:5 or SEQ ID NO:4, respectively. In some embodiments, variants of dCas9 or Cas9 nickase (e.g., variants of SEQ ID NO:5 and SEQ ID NO:4, respectively) are provided having amino acid sequences which are shorter, or longer than SEQ ID NO:5 or SEQ ID NO:4, by about 5 amino acids, by about 10 amino acids, by about 15 amino acids, by about 20 amino acids, by about 25 amino acids, by about 30 amino acids, by about 40 amino acids, by about 50 amino acids, by about 75 amino acids, by about 100 amino acids or more.
Cas9; nucleotide (Streptococcus pyogenes)

ATGGATAGTAAATCTCATGCTATGCTACTGCCAACAAATAGCTGCGTGGCGCCGATTATGATCGATGATATTATGACG
TGGCGCTAAGTTTACGTGCGCAGGAGATTGTAAGATAGATGCTAGACTGCTAGAAACACATCGCGACTGCTAGAAACTG
GAGAAGAAGTACGTGAGTCGAAATGATGCTAGACTGCTTGAGTAACTACACGAT

GGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAGATTTA
CTCCAAAAATACATGGGCGACTGCTATAGTACTGCTCGTTTACAGACTGTAAAAATTGTTGATGAACTGGTCAAAGTAA
AGTTTGACATTTAAAGAAGATATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGATTGCT

AAGGTCAAATATGTTACTGAGGGAATGCGAAAACCAGCATTTCTTTCAGGTAACAGAAGAAAGCCATTGTTGATA

CTTTTCCAGATATCTCATAGAAATAGTGAATAACTAAGCCTGCCATCTACGCTCAATTATAGCTTAAACTACCTAAATAT

AACTTAGCTGGCAGTCCTGCTATTAAAAAGGTATTTTACAGACTGTAAAAATTGTTGATGAACTGGTCAAAGTAA
AGTTTGACATTTAAAGAAGATATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGATTGCT

GAGAAGAAGTACGTGAGTCGAAATGATGCTAGACTGCTTGAGTAACTACACGAT

GGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAGATTTA
CTCCAAAAATACATGGGCGACTGCTATAGTACTGCTCGTTTACAGACTGTAAAAATTGTTGATGAACTGGTCAAAGTAA
AGTTTGACATTTAAAGAAGATATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGATTGCT

AAGGTCAAATATGTTACTGAGGGAATGCGAAAACCAGCATTTCTTTCAGGTAACAGAAGAAAGCCATTGTTGATA

CTTTTCCAGATATCTCATAGAAATAGTGAATAACTAAGCCTGCCATCTACGCTCAATTATAGCTTAAACTACCTAAATAT

AACTTAGCTGGCAGTCCTGCTATTAAAAAGGTATTTTACAGACTGTAAAAATTGTTGATGAACTGGTCAAAGTAA
AGTTTGACATTTAAAGAAGATATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGATTGCT

GAGAAGAAGTACGTGAGTCGAAATGATGCTAGACTGCTTGAGTAACTACACGAT

GGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAGATTTA
CTCCAAAAATACATGGGCGACTGCTATAGTACTGCTCGTTTACAGACTGTAAAAATTGTTGATGAACTGGTCAAAGTAA
AGTTTGACATTTAAAGAAGATATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGATTGCT

AAGGTCAAATATGTTACTGAGGGAATGCGAAAACCAGCATTTCTTTCAGGTAACAGAAGAAAGCCATTGTTGATA

CTTTTCCAGATATCTCATAGAAATAGTGAATAACTAAGCCTGCCATCTACGCTCAATTATAGCTTAAACTACCTAAATAT

AACTTAGCTGGCAGTCCTGCTATTAAAAAGGTATTTTACAGACTGTAAAAATTGTTGATGAACTGGTCAAAGTAA
AGTTTGACATTTAAAGAAGATATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGATTGCT

GAGAAGAAGTACGTGAGTCGAAATGATGCTAGACTGCTTGAGTAACTACACGAT
Cas9 (human codon optimized)

ATGGGACTAAAGTATTTGTTTGATGAGCAGTCGACGTATATTGCTACTAATATATCGGCGCTGTGAGTGTGTGTTGAGGAGATAGGCAAGGCTACACCGGAAATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAATTCCAGGACGAGGATTTTTATCCGTTCCTCAAAGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTACCTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGGAGGGTGATCTAAATGTGGGACCCCTGGCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCAATTTGGAAGAGTCCTTCCTTGCAAGAGGACAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTCGATTAAAAAGAATCTTATCGGTGCC

CCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGCATTCAAGGATAACGAACAGAAGCAACTTTTTGTTGAGCAGCACAAACATTATCTCGACGAAATCATAGAGCAAATTTCGGAAAAGCAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAATTCCAGGACGAGGATTTTTATCCGTTCCTCAAAGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTACCTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGGAGGGTGATCTAAATGTGGGACCCCTGGCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCAATTTGGAAGAGTCCTTCCTTGCAAGAGGACAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTCGATTAAAAAGAATCTTATCGGTGCC

AATCGCAGAATACCACTAATTGGAAGAAGTGGAAAATGGAGGAGAAGAAATATTGGAAGAAGAAGAAGAAGAAGAAGCTATTTTTCACAGTGTACAAT

TATTTCACAGTGTACAAT

TATTTCACAGTGTACAAT

GATGAGCTACTTGGAAGAGTCCTTCCTTGCAATTTTCATTGAGGGTACGCTACGCTGCAACG

ATCGCAGAATACCACTAATTGGAAGAAGTGGAAAATGGAGGAGAAGAAATATTGGAAGAAGAAGAAGAAGAAGAAGCTATTTTTCACAGTGTACAAT
Cas9; amino acid (Streptococcus pyogenes)

MDKY S IGLDITNSGVAVIDYEVKPSKFKVGLNDTRHS I KKNLIGALLFDGASTEAAETRLKRTARRRYTRR KNR ICYIQEFNSMEAKVDSSFFHRLEESLFLEDDEKKHERHP IFGNJIDEAYEHWKPY IYHLRRKVLVDSDKAD LRLYLYLALHMKFGRFHPLEGIDLNPSDNSVDFKFLQVLQTYNQLFEEENPINASGVDKAIALSRLKSRRLLEN IAQLPGEKKNFLGNIASIILSLGTPNFKSNFDLAE DAKLLSKLDVTDDDDLNLLAQI GIRQADYLFALKANLSLAI LSSDLIRVNETITEKPLASMIRKDEHHDQLTLLKLALVRQLEPVKYEKEEIPFDQSNGYAGYIDGASGEEYKFK IPIKLEMDGTELVLKINREDLLLKQRQDTDNS IPHQILHGLHELPALQRQDFEPFLLRDNKREI KELI TRF YVGPLARGNSRFMWTKRKSETITTPNWEEVEVDVKGASONFERMNTDFDNLPEKVLHSLYFELYFTYVNEILT KVKYTEGMRKPAFLSEQGKIAVDLLFLKNTIKVQVLKQKEDYFKKIECEDFSVE SIQVEDRNAFLSTHYDLDKI IJKDFKLDNENIEEELDVLTLTEFLDMRMRKLQQDKPIREQAENII HILFNTNLGAPAAF YKFDITT DRRKRTYSTKEVLDATLHIG SITGLYETRIDSQLDGD (SEQ ID NO:2)

(single underline: HNH domain; double underline: RuvC domain)

dCas9 (DIOA and H840A)

MDKY S IGLAIGTNSGVAVITDEYEVKPSKFKVGLNDTRHS I KKNLIGALLFDGASTEAAETRLKRTARRRYTRR KNR ICYIQE IFNSMAKVDSSFFHRLEESLFLEDDEKKHERHP IFGNJIDEAYEHWKPY IYHLRRKVLVDSDKAD LRLYLYLALHMKFGRFHPLEGIDLNPSDNSVDFKFLQVLQTYNQLFEEENPINASGVDKAIALSRLKSRRLLEN IAQLPGEKKNFLGNIASIILSLGTPNFKSNFDLAE DAKLLSKLDVTDDDDLNLLAQI GIRQADYLFALKANLSLAI LSSDLIRVNETITEKPLASMIRKDEHHDQLTLLKLALVRQLEPVKYEKEEIPFDQSNGYAGYIDGASGEEYKFK IPIKLEMDGTELVLKINREDLLLKQRQDTDNS IPHQILHGLHELPALQRQDFEPFLLRDNKREI KELI TRF YVGPLARGNSRFMWTKRKSETITTPNWEEVEVDVKGASONFERMNTDFDNLPEKVLHSLYFELYFTYVNEILT KVKYTEGMRKPAFLSEQGKIAVDLLFLKNTIKVQVLKQKEDYFKKIECEDFSVE SIQVEDRNAFLSTHYDLDKI IJKDFKLDNENIEEELDVLTLTEFLDMRMRKLQQDKPIREQAENII HILFNTNLGAPAAF YKFDITT DRRKRTYSTKEVLDATLHIG SITGLYETRIDSQLDGD (SEQ ID NO:2)

(single underline: HNH domain; double underline: RuvC domain)

Cas9 nickase (DIO A)(amino acid sequence)

MDKY S IGLAIGTNSGVAVITDEYEVKPSKFKVGLNDTRHS I KKNLIGALLFDGASTEAAETRLKRTARRRYTRR KNR ICYIQE IFNSMAKVDSSFFHRLEESLFLEDDEKKHERHP IFGNJIDEAYEHWKPY IYHLRRKVLVDSDKAD LRLYLYLALHMKFGRFHPLEGIDLNPSDNSVDFKFLQVLQTYNQLFEEENPINASGVDKAIALSRLKSRRLLEN IAQLPGEKKNFLGNIASIILSLGTPNFKSNFDLAE DAKLLSKLDVTDDDDLNLLAQI GIRQADYLFALKANLSLAI LSSDLIRVNETITEKPLASMIRKDEHHDQLTLLKLALVRQLEPVKYEKEEIPFDQSNGYAGYIDGASGEEYKFK IPIKLEMDGTELVLKINREDLLLKQRQDTDNS IPHQILHGLHELPALQRQDFEPFLLRDNKREI KELI TRF YVGPLARGNSRFMWTKRKSETITTPNWEEVEVDVKGASONFERMNTDFDNLPEKVLHSLYFELYFTYVNEILT KVKYTEGMRKPAFLSEQGKIAVDLLFLKNTIKVQVLKQKEDYFKKIECEDFSVE SIQVEDRNAFLSTHYDLDKI IJKDFKLDNENIEEELDVLTLTEFLDMRMRKLQQDKPIREQAENII HILFNTNLGAPAAF YKFDITT DRRKRTYSTKEVLDATLHIG SITGLYETRIDSQLDGD (SEQ ID NO:2)

(single underline: HNH domain; double underline: RuvC domain)
Cas9 variants are provided comprising an intein (e.g., a ligand-dependent intein) inserted within the Cas9 sequence and may be referred to as small-molecule-controlled Cas9 or ligand-dependent Cas9. In some embodiments, the intein is inserted into any location (e.g., at any amino acid position) in Cas9. In some embodiments, the inserted intein sequence replaces one or more amino acids in Cas9. For example, in some embodiments the inserted intein sequence replaces any cysteine, any alanine, any threonine, or any serine in Cas9 or a Cas9 variant such as dCas9 or Cas9 nickase. In some embodiments the inserted intein sequence replaces Cys80, Ala127, Thr146, Ser219, Thr333, Thr519, Cys574, Thr622, Ser701, Ala728, Thr995, Ser1006, Ser1154, Ser1159, or Ser1274 in Cas9 (SEQ ID NO:2), dCas9 (SEQ ID NO:5), or Cas9 nickase (SEQ ID NO:4).

The terms "conjugating," "conjugated," and "conjugation" refer to an association of two entities, for example, of two molecules such as two proteins, two domains (e.g., a binding domain and a cleavage domain), or a protein and an agent, e.g., a ligand binding domain and a small molecule. In some embodiments, the association is between a protein, RNA-programmable nuclease) and a nucleic acid (e.g., a guide RNA). The association can be, for example, by a direct or indirect (e.g., via a linker) covalent linkage. In some aspects, the association is between two or more proteins, for example, an RNA-programmable nuclease (e.g., Cas9) and an intein protein. In some embodiments, the association is covalent. In some embodiments, two molecules are conjugated via a linker connecting both molecules.

The term "consensus sequence," as used herein in the context of nucleic acid sequences, refers to a calculated sequence representing the most frequent nucleotide residues found at each position in a plurality of similar sequences. Typically, it represents the results of a multiple sequence alignments in which related sequences are compared to each other and similar sequence motifs are calculated. Methods and software for determining a consensus sequence are known in the art (See, e.g., JalCiew (jalview.org); and UGENE; Okonechnikov, K.; Golosova, O.; Fursov, M.; the UGENE team. "Unipro UGENE: a unified bioinformatics toolkit". Bioinformatics. 2012; doi:10.1093/bioinformatics/bts091).
The term "deaminase" refers to an enzyme that catalyzes a deamination reaction. In some embodiments, the deaminase is a cytidine deaminase, catalyzing the hydrolytic deamination of cytidine or deoxycytidine to uracil or deoxyuracil, respectively.

The term "effective amount," as used herein, refers to an amount of a biologically active agent (e.g., a ligand-dependent Cas9) that is sufficient to elicit a desired biological response. For example, in some embodiments, an effective amount of a nuclease may refer to the amount of the nuclease that is sufficient to induce cleavage of a desired target site-specifically bound and cleaved by the nuclease, preferably with minimal or no off-target cleavage. In some embodiments, an effective amount of another ligand-dependent Cas9 protein having other nucleic acid modifying activities may refer to the amount of the protein that is sufficient to induce the nucleic acid modification. As will be appreciated by the skilled artisan, the effective amount of an agent, e.g., a ligand-dependent nuclease, deaminase, recombinase, nickase, or a hybrid protein, a fusion protein, a protein dimer, a complex of a protein (or protein dimer) and a polynucleotide, or a polynucleotide, may vary depending on various factors as, for example, on the desired biological response, the specific allele, genome, target site, cell, or tissue being targeted, and the agent being used.

The term "engineered," as used herein, refers to a nucleic acid molecule, a protein molecule, complex, substance, or entity that has been designed, produced, prepared, synthesized, and/or manufactured by a human. Accordingly, an engineered product is a product that does not occur in nature.

The term "epigenetic modifier," as used herein, refers to a protein or catalytic domain thereof having enzymatic activity that results in the epigenetic modification of DNA, for example, chromosomal DNA. Epigenetic modifications include, but are not limited to, DNA methylation and demethylation; histone modifications including methylation and demethylation (e.g., mono-, di- and tri-methylation), histone acetylation and deacetylation, as well as histone ubiquitylation, phosphorylation, and sumoylation.

The term "extein," as used herein, refers to an polypeptide sequence that is flanked by an intein and is ligated to another extein during the process of protein splicing to form a mature, spliced protein. Typically, an intein is flanked by two extein sequences that are ligated together when the intein catalyzes its own excision. Exteins, accordingly, are the protein analog to exons found in mRNA. For example, a polypeptide comprising an intein may be of the structure extein(N) - intein - extein(C). After excision of the intein and splicing of the two exteins, the resulting structures are extein(N) - extein(C) and a free intein.
The term "hybrid protein," as used herein, refers to a protein that comprises the amino acid sequence of a target protein (e.g., a Cas9 protein) and, embedded in that amino acid sequence, a ligand-dependent intein as described herein. Accordingly, a hybrid protein generally comprises the structure: target protein(N) - intein - target protein(C). Typically, a hybrid protein comprises a Cas9 protein (e.g., Cas9, Cas9 variants such as dCas9, fragments of Cas9 or Cas9 variants, etc.) and a ligand-dependent intein. In some embodiments, a hybrid protein is encoded by a recombinant nucleic acid, in which a nucleic acid sequence encoding an intein is inserted in frame into a nucleic acid sequence encoding a target protein. In certain embodiments, the target protein exhibits a desired activity or property that is absent or reduced in the hybrid protein. In some embodiments, excision of the intein from the hybrid protein results in a restoration of the desired activity or property in the mature, spliced target protein. Non-limiting examples of desired activities or properties of target proteins are binding activities, enzymatic activities (e.g., nuclease activities, gene editing activities, deaminase activities, recombinase activities), reporter activities (e.g., fluorescent activity), therapeutic activities, size, charge, hydrophobicity, hydrophilicity, or 3D-structure. In some embodiments, excision of the intein from a hybrid protein results in a mature, spliced target protein that exhibits the same or similar levels of a desired activity as the native target protein. A hybrid protein may be created from any target protein by embedding an intein sequence into the amino acid sequence of the target protein, for example, by generating a recombinant, hybrid protein-encoding nucleic acid molecule and subsequent transcription and translation, or by protein synthesis methods known to those of skill in the art.

The term "intein," as used herein, refers to an amino acid sequence that is able to excise itself from a protein and to rejoin the remaining protein segments (the exteins) via a peptide bond in a process termed protein splicing. Inteins are analogous to the introns found in mRNA. Many naturally occurring and engineered inteins and hybrid proteins comprising such inteins are known to those of skill in the art, and the mechanism of protein splicing has been the subject of extensive research. As a result, methods for the generation of hybrid proteins from naturally occurring and engineered inteins are well known to the skilled artisan. For an overview, see pages 1-10, 193-207, 211-229, 233-252, and 325-341 of Gross, Belfort, Derbyshire, Stoddard, and Wood (Eds.) Homing Endonucleases and Inteins Springer Verlag Heidelberg, 2005; ISBN 9783540251064; the contents of which are incorporated herein by reference for disclosure of inteins and methods of generating hybrid proteins comprising natural or engineered inteins. As will be apparent to those of skill in the art, an intein may
catalyze protein splicing in a variety of extein contexts. Accordingly, an intein can be introduced into virtually any target protein sequence to create a desired hybrid protein, and the invention is not limited in the choice of target proteins.

[0045] The term "intein domain," as used herein, refers to the amino acid sequence of an intein that is essential for self-excision and extein ligation. For example, in some inteins, the entire intein amino acid sequence, or part(s) thereof, may constitute the intein domain, while in ligand-dependent inteins, the ligand-binding domain is typically embedded into the intein domain, resulting in the structure: intein domain (N) - ligand-binding domain - intein domain (C).

[0046] The term "ligand binding domain," as used herein, refers to a peptide or protein domain that binds a ligand. A ligand binding domain may be a naturally occurring domain or an engineered domain. Examples of ligand-binding domains referred to herein are the ligand binding domain of a native estrogen receptor, e.g., the ligand-binding domain of the native human estrogen receptor, and engineered, evolved, or mutated derivatives thereof. Other suitable ligand binding domains include the human thyroid hormone receptor (see, e.g., Skretas et al., "Regulation of protein activity with small-molecule-controlled inteins." Protein Sci. 2005; 14, 523-532) and members of the ribose-binding protein family (see, e.g., Bjorkman et al., "Multiple open forms of ribose-binding protein trace the path of its conformational change." J Mol Biol. 1998 12;279(3):651-64). Typically, a ligand-binding domain useful in the context of ligand-dependent inteins, as provided herein, exhibits a specific three-dimensional structure in the absence of the ligand, which inhibits intein self-excision, and undergoes a conformational change upon binding of the ligand, which promotes intein self-excision. Some of the ligand-dependent inteins provided herein comprise a ligand-binding domain derived from the estrogen receptor that can bind 4-HT and other estrogen-receptor ligands, e.g., ligands described in more detail elsewhere herein, and undergo a conformational change upon binding of the ligand. An appropriate ligand may be any chemical compound that binds the ligand-binding domain and induces a desired conformational change. In some embodiments, an appropriate ligand is a molecule that is bound by the ligand-binding domain with high specificity and affinity. In some embodiments, the ligand is a small molecule. In some embodiments, the ligand is a molecule that does not naturally occur in the context (e.g., in a cell or tissue) that a ligand-dependent intein is used in. For example, in some embodiments, the ligand-binding domain is a ligand-binding domain derived from an estrogen receptor, and the ligand is tamoxifen, or a derivative or analog thereof (e.g., 4-hydroxytamoxifen, 4-HT).
The term "ligand-dependent intein," as used herein refers to an intein that comprises a ligand-binding domain. Typically, the ligand-binding domain is inserted into the amino acid sequence of the intein, resulting in a structure intein (N) - ligand-binding domain - intein (C). Typically, ligand-dependent inteins exhibit no or only minimal protein splicing activity in the absence of an appropriate ligand, and a marked increase of protein splicing activity in the presence of the ligand. In some embodiments, the ligand-dependent intein does not exhibit observable splicing activity in the absence of ligand but does exhibit splicing activity in the presence of the ligand. In some embodiments, the ligand-dependent intein exhibits an observable protein splicing activity in the absence of the ligand, and a protein splicing activity in the presence of an appropriate ligand that is at least 5 times, at least 10 times, at least 50 times, at least 100 times, at least 150 times, at least 200 times, at least 250 times, at least 500 times, at least 1000 times, at least 1500 times, at least 2000 times, at least 2500 times, at least 5000 times, at least 10000 times, at least 50000 times, at least 1000000 times greater than the activity observed in the absence of the ligand. In some embodiments, the increase in activity is dose dependent over at least 1 order of magnitude, at least 2 orders of magnitude, at least 3 orders of magnitude, at least 4 orders of magnitude, or at least 5 orders of magnitude, allowing for fine-tuning of intein activity by adjusting the concentration of the ligand. Suitable ligand-dependent inteins are known in the art, and in include those provided below and those described in published U.S. Patent Application U.S. 2014/006571 I Al; Mootz et al., "Protein splicing triggered by a small molecule." J. Am. Chem. Soc. 2002; 124, 9044-9045; Mootz et al., "Conditional protein splicing: a new tool to control protein structure and function in vitro and in vivo." J. Am. Chem. Soc. 2003; 125, 10561-10569; Buskirk et al, Proc. Natl. Acad. Sci. USA. 2004; 101, 10505-10510; Skretas & Wood, "Regulation of protein activity with small-molecule-controlled inteins." Protein Sci. 2005; 14, 523-532; Schwartz, et al., "Post-translational enzyme activation in an animal via optimized conditional protein splicing." Nat. Chem. Biol. 2007; 3, 50-54; Peck et al., Chem. Biol. 2001; 18 (5), 619-630; the entire contents of each are hereby incorporated by reference.

2-4 intein:
CLAEGRIFDPTGTTHREDVVDGRKP IHVVAAKDGTLARPVSVSWFDQGTRDVIGLRIAGGAIVWATPDHKV LTETVWRAAGELRKGDVAGPGSGSNSLALSITADQMSVSLLDAEPP LLYEYDPTSPEASESMGLLLTNADRE LVHMINWAKRPGFVDLTLHDQAHLECAWLE LMIGLVWRSMEHPKLFAPLNLDDLDRNQKCVEGMVE IFDML LATSRRFMMLQGEFVCVLKS 11 LLNSGVYTLS STLKSEKDHIRALDKI TDTL IHLMAKAGTLQQHQR LAQLLI LSHIRHMSNKGMEHLYSMKYNNVPLYDDLLEMLDAHLRIAGGSGASRVQAFADALDDKFLHDMLAEE LRYSVIREVLPRARRTDFDLEVVELHTIVAEGVVHNC (SEQ ID NO:7)
3-2 intein:
CLAEGTRIFDPVTGITHRIEDVDGRKPIHVVAAKDGTLLARPVSVWFDQGTRDVIGLRIAGGAIVWATPDHKV
LTYEYWRAAGELRGKDRVAGPGGSNLSLLSLTDQVMSALLDAEPP1LYSEYDTPSFSEASMMGLTTNLADRE
LVHMINWAKRPGVFDLTLHDQHALLECWLIEIMILGVRSMEHPKLLFAPNLLLDRNQGKCVEGMVEIFDM
LATSSRFMNLLQGEFVCLKS ILLNSGVYFTFLSSTLKSLEEKDHIRALDKITDTLIHMAMALKLLQQHR
LQALLLLISHIRHMSKGMEHLYSMKYVNVPILDLLEMMDAHLRHAGGGSASRVQAFADALDDKFLHDMLAE
LRYSVIREVLTPTRARTFDLEVEELHTLVAEGVVVHNC (SEQ ID NO:8)

30R3-1 intein:
CLAEGTRIFDPVTGITHRIEDVDGRKPIHVVAAKDGTLLARPVSVWFDQGTRDVIGLRIAGGAIVWATPDHKV
LTYEYWRAAGELRGKDRVAGPGGSNLSLLSLTDQVMSALLDAEPP1LYSEYDTPSFSEASMMGLTTNLADRE
LVHMINWAKRPGVFDLTLHDQHALLECWLIEIMILGVRSMEHPKLLFAPNLLLDRNQGKCVEGMVEIFDM
LATSSRFMNLLQGEFVCLKS ILLNSGVYFTFLSSTLKSLEEKDHIRALDKITDTLIHMAMALKLLQQHR
LQALLLLISHIRHMSKGMEHLYSMKYVNVPILDLLEMMDAHLRHAGGGSASRVQAFADALDDKFLHDMLAE
LRYSVIREVLTPTRARTFDLEVEELHTLVAEGVVVHNC (SEQ ID NO:9)

30R3-2 intein:
CLAEGTRIFDPVTGITHRIEDVDGRKPIHVVAAKDGTLLARPVSVWFDQGTRDVIGLRIAGGAIVWATPDHKV
LTYEYWRAAGELRGKDRVAGPGGSNLSLLSLTDQVMSALLDAEPP1LYSEYDTPSFSEASMMGLTTNLADRE
LVHMINWAKRPGVFDLTLHDQHALLECWLIEIMILGVRSMEHPKLLFAPNLLLDRNQGKCVEGMVEIFDM
LATSSRFMNLLQGEFVCLKS ILLNSGVYFTFLSSTLKSLEEKDHIRALDKITDTLIHMAMALKLLQQHR
LQALLLLISHIRHMSKGMEHLYSMKYVNVPILDLLEMMDAHLRHAGGGSASRVQAFADALDDKFLHDMLAE
LRYSVIREVLTPTRARTFDLEVEELHTLVAEGVVVHNC (SEQ ID NO:10)

30R3-3 intein:
CLAEGTRIFDPVTGITHRIEDVDGRKPIHVVAAKDGTLLARPVSVWFDQGTRDVIGLRIAGGAIVWATPDHKV
LTYEYWRAAGELRGKDRVAGPGGSNLSLLSLTDQVMSALLDAEPP1LYSEYDTPSFSEASMMGLTTNLADRE
LVHMINWAKRPGVFDLTLHDQHALLECWLIEIMILGVRSMEHPKLLFAPNLLLDRNQGKCVEGMVEIFDM
LATSSRFMNLLQGEFVCLKS ILLNSGVYFTFLSSTLKSLEEKDHIRALDKITDTLIHMAMALKLLQQHR
LQALLLLISHIRHMSKGMEHLYSMKYVNVPILDLLEMMDAHLRHAGGGSASRVQAFADALDDKFLHDMLAE
LRYSVIREVLTPTRARTFDLEVEELHTLVAEGVVVHNC (SEQ ID NO:11)

37R3-1 intein:
CLAEGTRIFDPVTGITHRIEDVDGRKPIHVVAAKDGTLLARPVSVWFDQGTRDVIGLRIAGGAIVWATPDHKV
LTYEYWRAAGELRGKDRVAGPGGSNLSLLSLTDQVMSALLDAEPP1LYSEYDTPSFSEASMMGLTTNLADRE
LVHMINWAKRPGVFDLTLHDQHALLECWLIEIMILGVRSMEHPKLLFAPNLLLDRNQGKCVEGMVEIFDM
LATSSRFMNLLQGEFVCLKS ILLNSGVYFTFLSSTLKSLEEKDHIRALDKITDTLIHMAMALKLLQQHR
LQALLLLISHIRHMSKGMEHLYSMKYVNVPILDLLEMMDAHLRHAGGGSASRVQAFADALDDKFLHDMLAE
LRYSVIREVLTPTRARTFDLEVEELHTLVAEGVVVHNC (SEQ ID NO:12)

37R3-2 intein:
CLAEGTRIFDPVTGITHRIEDVDGRKPIHVVAAKDGTLLARPVSVWFDQGTRDVIGLRIAGGAIVWATPDHKV
LTYEYWRAAGELRGKDRVAGPGGSNLSLLSLTDQVMSALLDAEPP1LYSEYDTPSFSEASMMGLTTNLADRE
LVHMINWAKRPGVFDLTLHDQHALLECWLIEIMILGVRSMEHPKLLFAPNLLLDRNQGKCVEGMVEIFDM
LATSSRFMNLLQGEFVCLKS ILLNSGVYFTFLSSTLKSLEEKDHIRALDKITDTLIHMAMALKLLQQHR
LQALLLLISHIRHMSKGMEHLYSMKYVNVPILDLLEMMDAHLRHAGGGSASRVQAFADALDDKFLHDMLAE
LRYSVIREVLTPTRARTFDLEVEELHTLVAEGVVVHNC (SEQ ID NO:13)

37R3-3 intein:
CLAEGTRIFDPVTGITHRIEDVDGRKPIHVVAAKDGTLLARPVSVWFDQGTRDVIGLRIAGGAIVWATPDHKV
LTYEYWRAAGELRGKDRVAGPGGSNLSLLSLTDQVMSALLDAEPP1LYSEYDTPSFSEASMMGLTTNLADRE
LVHMINWAKRPGVFDLTLHDQHALLECWLIEIMILGVRSMEHPKLLFAPNLLLDRNQGKCVEGMVEIFDM
LATSSRFMNLLQGEFVCLKS ILLNSGVYFTFLSSTLKSLEEKDHIRALDKITDTLIHMAMALKLLQQHR
LQALLLLISHIRHMSKGMEHLYSMKYVNVPILDLLEMMDAHLRHAGGGSASRVQAFADALDDKFLHDMLAE
LRYSVIREVLTPTRARTFDLEVEELHTLVAEGVVVHNC (SEQ ID NO:14)
The term "linker," as used herein, refers to a chemical group or a molecule linking two adjacent molecules or moieties, e.g., two polypeptides. Typically, the linker is positioned between, or flanked by, two groups, molecules, or other moieties and connected to each one via a covalent bond, thus connecting the two. In some embodiments, the linker is an amino acid linker. In some embodiments, the amino acid linker comprises at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, or at least 30 amino acids. In some embodiments, the linker is a divalent organic molecule, group, polymer, or chemical moiety. In some embodiments, the peptide linker comprises repeats of the tri-peptide Gly-Gly-Ser, e.g., comprising the sequence (GGS)n, wherein n represents at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more repeats. In some embodiments, the linker comprises the sequence (GGS)6 (SEQ ID NO: 15). In some embodiments, the peptide linker is the 16 residue "XTEN" linker, or a variant thereof (See, e.g., Schellenberger et al. A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. Nat. Biotechnol. 27, 1186-1190 (2009)). In some embodiments, the XTEN linker comprises the sequence SGSETPGTSESATPES (SEQ ID NO: 16), SGSETPGTSESASA (SEQ ID NO: 17), or SGSETPGTSESATPEGGSGGS (SEQ ID NO: 18). In some embodiments, the peptide linker is one or more selected from VPFLLEPDPNDINGKTC (SEQ ID NO: 19), GSAGSAAGSGEF (SEQ ID NO:20), SIVAQLSRPDPDA (SEQ ID NO:21), MKIEEQLPSA (SEQ ID NO:22), VRHKLRKVGS (SEQ ID NO:23), GHGTGSTGSGSS (SEQ ID NO:24), MSRPDPDA (SEQ ID NO:25); or GGSM (SEQ ID NO:26).

The term "mutation," as used herein, refers to a substitution of a residue within a sequence, e.g., a nucleic acid or amino acid sequence, with another residue, or a deletion or insertion of one or more residues within a sequence. Mutations are typically described herein by identifying the original residue followed by the position of the residue within the sequence and by the identity of the newly substituted residue. Methods for making the amino acid substitutions (mutations) provided herein are known in the art and are provided by, for example, Green and Sambrook, Molecular Cloning: A Laboratory Manual (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)).

The term "nuclease," as used herein, refers to an agent, for example, a protein, capable of cleaving a phosphodiester bond connecting two nucleotide residues in a nucleic
acid molecule. In some embodiments, "nuclease" refers to a protein having an inactive DNA cleavage domain, such that the nuclease is incapable of cleaving a phosphodiester bond. In some embodiments, the nuclease is a protein, e.g., an enzyme that can bind a nucleic acid molecule and cleave a phosphodiester bond connecting nucleotide residues within the nucleic acid molecule. A nuclease may be an endonuclease, cleaving a phosphodiester bonds within a polynucleotide chain, or an exonuclease, cleaving a phosphodiester bond at the end of the polynucleotide chain. In some embodiments, a nuclease is a site-specific nuclease, binding and/or cleaving a specific phosphodiester bond within a specific nucleotide sequence, which is also referred to herein as the "recognition sequence," the "nuclease target site," or the "target site." In some embodiments, a nuclease is an RNA-guided (i.e., RNA-programmable) nuclease, which is associated with (e.g., binds to) an RNA (e.g., a guide RNA, "gRNA") having a sequence that complements a target site, thereby providing the sequence specificity of the nuclease. In some embodiments, a nuclease recognizes a single stranded target site. In some embodiments, a nuclease recognizes a double-stranded target site, for example, a double-stranded DNA target site. The target sites of many naturally occurring nucleases, for example, many naturally occurring DNA restriction nucleases, are well known to those of skill in the art. In many cases, a DNA nuclease, such as EcoRI, HindIII, or BamHI, recognize a palindromic, double-stranded DNA target site of 4 to 10 base pairs in length, and cut each of the two DNA strands at a specific position within the target site. Some endonucleases cut a double-stranded nucleic acid target site symmetrically, i.e., cutting both strands at the same position so that the ends comprise base-paired nucleotides, also referred to herein as blunt ends. Other endonucleases cut a double-stranded nucleic acid target site asymmetrically, i.e., cutting each strand at a different position so that the ends include unpaired nucleotides. Unpaired nucleotides at the end of a double-stranded DNA molecule are also referred to as "overhangs," e.g., as "5'-overhang" or as "3'-overhang," depending on whether the unpaired nucleotide(s) form(s) the 5' or the 3' end of the respective DNA strand. Double-stranded DNA molecule ends ending with unpaired nucleotide(s) are also referred to as sticky ends, as they can "stick to" other double-stranded DNA molecule ends comprising complementary unpaired nucleotide(s). A nuclease protein typically comprises a "binding domain" that mediates the interaction of the protein with the nucleic acid substrate, and also, in some cases, specifically binds to a target site, and a "cleavage domain" that catalyzes the cleavage of the phosphodiester bond within the nucleic acid backbone. In some embodiments, a nuclease protein can bind and cleave a nucleic acid molecule in a monomeric form. In some embodiments, a nuclease protein has to dimerize or multimerize in order to
cleave a target nucleic acid. Binding domains and cleavage domains of naturally occurring nuclease, as well as modular binding domains and cleavage domains that can be fused to create nucleases binding specific target sites, are well known to those of skill in the art. For example, the binding domain of RNA-programmable nucleases (e.g., Cas9), or a Cas9 protein having an inactive DNA cleavage domain (e.g., dCas9), can be used as a binding domain (e.g., that binds a gRNA to direct binding to a target site) to specifically bind a desired target site, and fused or conjugated to a cleavage domain, for example, the cleavage domain of FokI, to create an engineered nuclease cleaving the target site. In some embodiments, Cas9 fusion proteins provided herein comprise the cleavage domain of FokI, and are therefore referred to as "fCas9" proteins. In some embodiments, the cleavage domain of FokI, e.g., in a fCas9 protein corresponds to, or comprises in part or whole, the amino acid sequence (or variants thereof) set forth as SEQ ID NO:6, below. In some embodiments, variants or homologues of the FokI cleavage domain include any variant or homologue capable of dimerizing (e.g., as part of fCas9 fusion protein) with another FokI cleavage domain at a target site in a target nucleic acid, thereby resulting in cleavage of the target nucleic acid. In some embodiments, variants of the FokI cleavage domain (e.g., variants of SEQ ID NO:6) are provided which are at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to SEQ ID NO:6. In some embodiments, variants of the FokI cleavage domain (e.g., variants of SEQ ID NO:6) are provided having an amino acid sequence which is shorter, or longer than SEQ ID NO:6, by about 5 amino acids, by about 10 amino acids, by about 15 amino acids, by about 20 amino acids, by about 25 amino acids, by about 30 amino acids, by about 40 amino acids, by about 50 amino acids, by about 75 amino acids, by about 100 amino acids, or more.

Cleavage domain of FokI:

GSQLVKSEXLEEKKSELHRKLKYVPHEYIELIEIARNSTQDRIEMKMKEFLMKVYGKYRGKHLGSSRKPDGAIYTV
GSPIDYGIVDTKAYSGGYNLPQGADEMQRYVEENQTRNKHNIPNEWKVYPSSVTEFKPLEVSQHFKGNYKAIQLTRLNHITNCNGAVLSEELIUIGIKAGTLTLEEVRRKFNNGEINF (SEQ ID NO:6)

[0051] The terms "nucleic acid" and "nucleic acid molecule," as used herein, refer to a compound comprising a nucleobase and an acidic moiety, e.g., a nucleoside, a nucleotide, or a polymer of nucleotides. Typically, polymeric nucleic acids, e.g., nucleic acid molecules comprising three or more nucleotides are linear molecules, in which adjacent nucleotides are linked to each other via a phosphodiester linkage. In some embodiments, the term "nucleic
"acid" refers to individual nucleic acid residues (e.g. nucleotides and/or nucleosides). In some embodiments, "nucleic acid" refers to an oligonucleotide chain comprising three or more individual nucleotide residues. As used herein, the terms "oligonucleotide" and "polynucleotide" can be used interchangeably to refer to a polymer of nucleotides (e.g., a string of at least three nucleotides). In some embodiments, "nucleic acid" encompasses RNA as well as single and/or double-stranded DNA. Nucleic acids may be naturally occurring, for example, in the context of a genome, a transcript, an mRNA, tRNA, rRNA, siRNA, snRNA, a plasmid, cosmid, chromosome, chromatid, or other naturally occurring nucleic acid molecule. On the other hand, a nucleic acid molecule may be a non-naturally occurring molecule, e.g., a recombinant DNA or RNA, an artificial chromosome, an engineered genome, or fragment thereof, or a synthetic DNA, RNA, DNA/RNA hybrid, or including non-naturally occurring nucleotides or nucleosides. Furthermore, the terms "nucleic acid," "DNA," "RNA," and/or similar terms include nucleic acid analogs, i.e. analogs having other than a phosphodiester backbone. Nucleic acids can be purified from natural sources, produced using recombinant expression systems and optionally purified, chemically synthesized, etc. Where appropriate, e.g., in the case of chemically synthesized molecules, nucleic acids can comprise nucleoside analogs such as analogs having chemically modified bases or sugars, and backbone modifications. A nucleic acid sequence is presented in the 5' to 3' direction unless otherwise indicated. In some embodiments, a nucleic acid is or comprises natural nucleosides (e.g. adenosine, thymidine, guanosine, cytidine, uridine, deoxyadenosine, deoxymethylidine, deoxyguanosine, and deoxycytidine); nucleoside analogs (e.g., 2-aminoadenosine, 2-thiothymidine, inosine, pyrrolo-pyrimidine, 3-methyl adenosine, 5-methylcytidine, 2-aminoadenosine, C5-bromouridine, C5-fluorouridine, C5-iodouridine, C5-propynyl-uridine, C5-propynyl-cytidine, C5-methylcytidine, 2-aminoadenosine, 7-deazaadenosine, 7-deazaguanosine, 8-oxadenosine, 8-oxoguanosine, 0(6)-methylguanine, and 2-thiocytidine); chemically modified bases; biologically modified bases (e.g., methylated bases); intercalated bases; modified sugars (e.g., 2'-fluororibose, ribose, 2'-deoxyribose, arabinose, and hexose); and/or modified phosphate groups (e.g., phosphorothioates and 5'-N-phosphorimidate linkages).

[0052] The term "pharmaceutical composition," as used herein, refers to a composition that can be administrated to a subject in the context of treatment of a disease or disorder. In some embodiments, a pharmaceutical composition comprises an active ingredient, e.g., a nuclease or a nucleic acid encoding a nuclease, and a pharmaceutically acceptable excipient.
The term "proliferative disease," as used herein, refers to any disease in which cell or tissue homeostasis is disturbed in that a cell or cell population exhibits an abnormally elevated proliferation rate. Proliferative diseases include hyperproliferative diseases, such as pre-neoplastic hyperplastic conditions and neoplastic diseases. Neoplastic diseases are characterized by an abnormal proliferation of cells and include both benign and malignant neoplasias. Malignant neoplasia is also referred to as cancer.

The terms "protein," "peptide," and "polypeptide" are used interchangeably herein and refer to a polymer of amino acid residues linked together by peptide (amide) bonds. The terms refer to a protein, peptide, or polypeptide of any size, structure, or function. Typically, a protein, peptide, or polypeptide will be at least three amino acids long. A protein, peptide, or polypeptide may refer to an individual protein or a collection of proteins. One or more of the amino acids in a protein, peptide, or polypeptide may be modified, for example, by the addition of a chemical entity such as a carbohydrate group, a hydroxyl group, a phosphate group, a farnesyl group, an isofarnesyl group, a fatty acid group, a linker for conjugation, functionalization, or other modification, etc. A protein, peptide, or polypeptide may also be a single molecule or may be a multi-molecular complex. A protein, peptide, or polypeptide may be just a fragment of a naturally occurring protein or peptide. A protein, peptide, or polypeptide may be naturally occurring, recombinant, or synthetic, or any combination thereof.

The term "fusion protein" as used herein refers to a hybrid polypeptide which comprises protein domains from at least two different proteins. One protein may be located at the amino-terminal (N-terminal) portion of the fusion protein or at the carboxy-terminal (C-terminal) protein thus forming an "amino-terminal fusion protein" or a "carboxy-terminal fusion protein," respectively. A protein may comprise different domains, for example, a nucleic acid binding domain (e.g., the gRNA binding domain of Cas9 that directs the binding of the protein to a target site or a dCas9 protein) and a nucleic acid cleavage domain(s). In some embodiments, a protein is in a complex with, or is in association with, a nucleic acid, e.g., DNA or RNA. Any of the proteins provided herein may be produced by any method known in the art. For example, the proteins provided herein may be produced via recombinant protein expression and purification, which is especially suited for fusion proteins comprising a peptide linker. Methods for recombinant protein expression and purification are well known and include those described by Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)), the entire contents of which are incorporated herein by reference.
The term "protein splicing," as used herein, refers to a process in which a sequence, an intein, is excised from within an amino acid sequence, and the remaining fragments of the amino acid sequence, the exteins, are ligated via an amide bond to form a continuous amino acid sequence.

The term "RNA-programmable nuclease" and "RNA-guided nuclease" are used interchangeably herein and refer to a nuclease that forms a complex with (e.g., binds or associates with) one or more RNAs that is not a target for cleavage. In some embodiments, an RNA-programmable nuclease, when in a complex with an RNA, may be referred to as a nuclease:RNA complex. Typically, the bound RNA(s) is referred to as a guide RNA (gRNA). gRNAs can exist as an association of two or more RNAs, or as a single RNA molecule. gRNAs that exist as a single RNA molecule may be referred to as single-guide RNAs (sgRNAs), though "gRNA" is used interchangeably to refer to guide RNAs that exist as either a single molecule or as a complex of two or more molecules. Typically, gRNAs that exist as single RNA species comprise at least two domains: (1) a domain that shares homology to a target nucleic acid and may direct binding of a Cas9 complex to the target; and (2) a domain that binds a Cas9 protein. In some embodiments, domain (2) corresponds to a sequence known as a tracrRNA and comprises a stem-loop structure. For example, in some embodiments, domain (2) is homologous to a tracrRNA as depicted in Figure 1E of Jinek et al., Science 337:816-821(2012), the entire contents of which is incorporated herein by reference. In some embodiments, domain 2 is at least 90%, at least 95%, at least 98%, or at least 99% identical to the tracrRNA as described by Jinek et al., Science 337:816-821(2012). The gRNA comprises a nucleotide sequence that complements a target site, which mediates binding of the nuclease:RNA complex to said target site and the sequence specificity of the nuclease:RNA complex. The sequence of a gRNA that binds a target nucleic acid can comprise any sequence that complements a region of the target and is suitable for a nuclease:RNA complex to bind. In some embodiments, the RNA-programmable nuclease is the (CRISPR-associated system) Cas9 endonuclease, for example, Cas9 (Csn1) from Streptococcus pyogenes (see, e.g., "Complete genome sequence of an M1 strain of Streptococcus pyogenes." Ferretti J.J., McShan W.M., Ajdíc D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L. expand/collapse author list McLaughlin R.E., Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Delcheva E., Chylinski K., Sharma CM., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607(2011);
and "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial
immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E.
Science 337:816-821(2012), the entire contents of each of which are incorporated herein by reference.

Because RNA-programmable nuclease (e.g., Cas9) use RNA:DNA hybridization to determine target DNA cleavage sites, these proteins are able to cleave, in principle, any sequence specified by the guide RNA. Methods of using RNA-programmable nucleases, such as Cas9, for site-specific cleavage (e.g., to modify a genome) are known in the art (see e.g., Cong et al. Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819-823 (2013); Mali et al. RNA-guided human genome engineering via Cas9. Science 339, 823-826 (2013); Hwang et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nature biotechnology 31, 227-229 (2013); Jinek et al. RNA-programmed genome editing in human cells. eLife 2, e00471 (2013); Dicarlo et al. Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic acids research (2013); Jiang et al. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nature biotechnology 31, 233-239 (2013); the entire contents of each of which are incorporated herein by reference).

The term "recombinase," as used herein, refers to a site-specific enzyme that mediates the recombination of DNA between recombinase recognition sequences, which results in the excision, integration, inversion, or exchange (e.g., translocation) of DNA fragments between the recombinase recognition sequences. Recombinases can be classified into two distinct families: serine recombinases (e.g., resolvases and invertases) and tyrosine recombinases (e.g., integrases). Examples of serine recombinases include, without limitation, Hin, Gin, Tn3, β-six, CinH, ParA, γØ, Bxbl, 4>C31, TP901, TGI, φBTI, R4, φR V 1, 4>FC1, MRU, A118, U153, and gp29. Examples of tyrosine recombinases include, without limitation, Cre, FLP, R, Lambda, HK101, HK022, and pSAM2. The serine and tyrosine recombinase names stem from the conserved nucleophilic amino acid residue that the recombinase uses to attack the DNA and which becomes covalently linked to the DNA during strand exchange. Recombinases have numerous applications, including the creation of gene knockouts/knock-ins and gene therapy applications. See, e.g., Brown et al., "Serine recombinases as tools for genome engineering." Methods. 2011;53(4):372-9; Hirano et al., "Site-specific recombinases as tools for heterologous gene integration." Appl. Microbiol. Biotechnol. 2011; 92(2):227-39; Chavez and Calos, "Therapeutic applications of the ΦC31 integrase system." Curr. Gene Ther. 2011;11(5):375-81; Turan and Bode, "Site-specific
recombinases: from tag-and-target- to tag-and-exchange-based genomic modifications."

FASEB J. 2011; 25(12):4088-107; Venken and Bellen, "Genome-wide manipulations of Drosophila melanogaster with transposons, Flp recombinase, and F3 integrase." Methods Mol. Biol. 2012; 859:203-28; Murphy, "Phage recombinases and their applications." Adv. Virus Res. 2012; 83:367-414; Zhang et al., "Conditional gene manipulation: Cre-ating a new biological era." J. Zhejiang Univ. Sci. B. 2012; 13(7):5 11-24; Karpenshif and Bernstein, "From yeast to mammals: recent advances in genetic control of homologous recombination." DNA Repair (Amst). 2012; 11:710:781-8; the entire contents of each are hereby incorporated by reference in their entirety. The recombinases provided herein are not meant to be exclusive examples of recombinases that can be used in embodiments of the invention. The methods and compositions of the invention can be expanded by mining databases for new orthogonal recombinases or designing synthetic recombinases with defined DNA specificities (See, e.g., Groth et al., "Phage integrases: biology and applications." J. Mol. Biol. 2004; 335, 667-678; Gordley et al., "Synthesis of programmable integrases." Proc. Natl. Acad. Sci. USA. 2009; 106, 5053-5058; the entire contents of each are hereby incorporated by reference in their entirety). Other examples of recombinases that are useful in the methods and compositions described herein are known to those of skill in the art, and any new recombinase that is discovered or generated is expected to be able to be used in the different embodiments of the invention. In some embodiments, the catalytic domains of a recombinase are fused to a nuclease-inactivated RNA-programmable nuclease (e.g., dCas9, or a fragment thereof), such that the recombinase domain does not comprise a nucleic acid binding domain or is unable to bind to a target nucleic acid (e.g., the recombinase domain is engineered such that it does not have specific DNA binding activity). Recombinases lacking DNA binding activity and methods for engineering such are known, and include those described by Klippel et al., "Isolation and characterisation of unusual gin mutants." EMBO J. 1988; 7:3983-3989; Burke et al., "Activating mutations of Tn3 resolvase marking interfaces important in recombination catalysis and its regulation. Mol Microbiol. 2004; 51: 937-948; Olorunniji et al., "Synapsis and catalysis by activated Tn3 resolvase mutants." Nucleic Acids Res. 2008; 36: 7181-7191; Rowland et al., "Regulatory mutations in Sin recombinase support a structure -based model of the synaptosome." Mol Microbiol. 2009; 74: 282-298; Akopian et al., "Chimeric recombinases with designed DNA sequence recognition." Proc Natl Acad Sci USA. 2003;100: 8688-8691; Gordley et al., "Evolution of programmable zinc finger-recombinases with activity in human cells. J Mol Biol. 2007; 367: 802-813; Gordley et al., "Synthesis of programmable integrases." Proc Natl Acad Sci USA. 2009;106: 5053-
5058; Arnold et al, "Mutants of TnJ resolvase which do not require accessory binding sites for recombination activity." EMBO J. 1999;18: 1407-1414; Gaj et al, "Structure-guided reprogramming of serine recombinase DNA sequence specificity." Proc Natl Acad Sci USA. 2011;108(2):498-503; and Proudfoot et al, "Zinc finger recombinases with adaptable DNA sequence specificity." PLoS One. 2011;6(4):e19537; the entire contents of each are hereby incorporated by reference. For example, serine recombinases of the resolvase-invertase group, e.g., Tn3 and γ8 resolvases and the Hin and Gin invertases, have modular structures with autonomous catalytic and DNA-binding domains (See, e.g., Grindley et al, "Mechanism of site-specific recombination." Ann Rev Biochem. 2006; 75: 567-605, the entire contents of which are incorporated by reference). The catalytic domains of these recombinases are thus amenable to being recombined with nuclease-inactivated RNA-programmable nucleases (e.g., dCas9, or a fragment thereof) as described herein, e.g., following the isolation of 'activated' recombinase mutants which do not require any accessory factors (e.g., DNA binding activities) (See, e.g., Klippel et al, "Isolation and characterisation of unusual gin mutants." EMBO J. 1988; 7: 3983-3989: Burke et al, "Activating mutations of Tn3 resolvase marking interfaces important in recombination catalysis and its regulation. Mol Microbiol. 2004; 51: 937-948; Olorunniji et al, "Synapsis and catalysis by activated Tn3 resolvase mutants." Nucleic Acids Res. 2008; 36: 7181-7191; Rowland et al, "Regulatory mutations in Sin recombinase support a structure-based model of the synaptosome." Mol Microbiol. 2009; 74: 282-298; Akopian et al, "Chimeric recombinases with designed DNA sequence recognition." Proc Natl Acad Sci USA. 2003;100: 8688-8691). Additionally, many other natural serine recombinases having an N-terminal catalytic domain and a C-terminal DNA binding domain are known (e.g., phiC31 integrase, TnpX transposase, IS607 transposase), and their catalytic domains can be co-opted to engineer programmable site-specific recombinases as described herein (See, e.g., Smith et al, "Diversity in the serine recombinases." Mol Microbiol. 2002;44: 299-307, the entire contents of which are incorporated by reference). Similarly, the core catalytic domains of tyrosine recombinases (e.g., Cre, λ integrase) are known, and can be similarly co-opted to engineer programmable site-specific recombinases as described herein (See, e.g., Guo et al, "Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse." Nature. 1997; 389:40-46; Hartung et al, "Cre mutants with altered DNA binding properties." J Biol Chem 1998; 273:22884-22891; Shaikh et al, "Chimeras of the Flp and Cre recombinases: Tests of the mode of cleavage by Flp and Cre. J Mol Biol. 2000; 302:27-48; Rongrong et al, "Effect of deletion mutation on the recombination activity of Cre recombinase." Acta Biochim Pol.

[0060] The term "recombine," or "recombination," in the context of a nucleic acid modification {e.g., a genomic modification}, is used to refer to the process by which two or more nucleic acid molecules, or two or more regions of a single nucleic acid molecule, are modified by the action of a recombinase protein {e.g., an inventive recombinase fusion protein provided herein}. Recombination can result in, inter alia, the insertion, inversion, excision, or translocation of nucleic acids, e.g., in or between one or more nucleic acid molecules.

[0061] The term "site-specific enzyme," as used herein, refers to any enzyme capable of binding a nucleic acid at a target site to mediate a modification of the nucleic acid. Typically, the site-specific enzymes provided herein comprise an intein {e.g., a ligand-dependent intein}. In some embodiments, the site-specific enzyme is unable to bind a target site prior to excision of the intein. In some embodiments, the site-specific enzyme is able to bind a target site prior to excision of the intein but remains enzymatically inactive {e.g., cannot cleave, recombine, edit, or otherwise modify a nucleic acid} until excision of the intein.

[0062] The term "small molecule," as used herein, refers to a non-peptidic, non-oligomeric organic compound either prepared in the laboratory or found in nature. Small molecules, as used herein, can refer to compounds that are "natural product-like", however, the term "small molecule" is not limited to "natural product-like" compounds. Rather, a small molecule is typically a non-polymeric, non-oligomeric molecule that is characterized in
that it contains several carbon-carbon bonds, and has a molecular weight of less than 2000g/mol, preferably less than 1500g/mol, although this characterization is not intended to be limiting for the purposes of the present invention. In certain embodiments, the ligand of a ligand-dependent inteins used in the present invention is a small molecule.

[0063] The term "subject," as used herein, refers to an individual organism. In some embodiments, the subject is a human. In some embodiments, the subject is a non-human mammal. In some embodiments, the subject is a non-human primate. In some embodiments, the subject is a rodent. In some embodiments, the subject is a sheep, a goat, a cattle, a cat, or a dog. In some embodiments, the subject is a vertebrate, an amphibian, a reptile, a fish, an insect, a fly, or a nematode. In some embodiments, the subject is a research animal. In some embodiments, the subject is genetically engineered, e.g., a genetically engineered non-human subject. The subject may be of either sex and at any stage of development.

[0064] The terms "target nucleic acid," and "target genome," as used herein in the context of nuclease, refer to a nucleic acid molecule or a genome, respectively, that comprises at least one target site of a given nuclease.

[0065] The term "target site," refers to a sequence within a nucleic acid molecule that is bound and (1) cleaved; (2) recombined; (3) edited; or (4) otherwise modified by a site-specific enzyme. In some embodiments, a target site refers to a "nuclease target site," which is a sequence within a nucleic acid molecule that is bound and cleaved by a nuclease. A target site may be single-stranded or double-stranded. In the context of RNA-guided (i.e., RNA-programmable) nuclease (e.g., a Cas9 protein, a Cas9 variant, fragments of Cas9 or fragments of Cas9 variants, etc.), a target site typically comprises a nucleotide sequence that is complementary to a gRNA of the RNA-guided nuclease, and a protospacer adjacent motif (PAM) at the 3' end adjacent to the gRNA-complementary sequence. For the RNA-guided nuclease Cas9, the target site may be, in some embodiments, 20 base pairs plus a 3 base pair PAM (e.g., NNN, wherein N represents any nucleotide). Typically, the first nucleotide of a PAM can be any nucleotide, while the two downstream nucleotides are specified depending on the specific RNA-guided nuclease. Exemplary target sites for RNA-guided nucleases, such as Cas9, are known to those of skill in the art and include, without limitation, NNG, NGN, NAG, and NGG, wherein N represents any nucleotide. In addition, Cas9 nucleases from different species (e.g., S. thermophilus instead of S. pyogenes) recognize a PAM that comprises the sequence: NGGNG. Additional PAM sequences are known, including, but not limited to, NNAGAAW and NAAR (see, e.g., Esvelt and Wang, Molecular Systems Biology, 9:641 (2013), the entire contents of which are incorporated herein by reference). For
example, the target site of an RNA-guided nuclease, such as, e.g., Cas9, may comprise the structure \([Nz][-PAM]\), where each \(N\) is, independently, any nucleotide, and \(z\) is an integer between 1 and 50. In some embodiments, \(z\) is at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, or at least 50. In some embodiments, \(z\) is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50. In some embodiments, \(Z\) is 20. In some embodiments, "target site" may also refer to a sequence within a nucleic acid molecule that is bound but not cleaved by a nuclease.

[0066] The terms "transcriptional activator" and "transcriptional repressor," refer to agents which activate and repress the transcription of a gene, respectively. Typically, such activators and repressors are proteins, e.g., as provided herein.

[0067] The terms "treatment," "treat," and "treating," refer to a clinical intervention aimed to reverse, alleviate, delay the onset of, or inhibit the progress of a disease or disorder, or one or more symptoms thereof, as described herein. As used herein, the terms "treatment," "treat," and "treating" refer to a clinical intervention aimed to reverse, alleviate, delay the onset of, or inhibit the progress of a disease or disorder, or one or more symptoms thereof, as described herein. In some embodiments, treatment may be administered after one or more symptoms have developed and/or after a disease has been diagnosed. In some embodiments, treatment may be administered in the absence of symptoms, e.g., to prevent or delay onset of a symptom or inhibit onset or progression of a disease. For example, treatment may be administered to a susceptible individual prior to the onset of symptoms (e.g., in light of a history of symptoms and/or in light of genetic or other susceptibility factors). Treatment may also be continued after symptoms have resolved, for example, to prevent or delay their recurrence.

[0068] The term "vector" refers to a polynucleotide comprising one or more recombinant polynucleotides of the present invention, e.g., those encoding or a Cas9 protein (e.g., a Cas9 protein comprising an intein) and/or a gRNA provided herein. Vectors include, but are not limited to, plasmids, viral vectors, cosmids, artificial chromosomes, and phagemids. The vector is one which is able to replicate in a host cell, and which is further characterized by one or more endonuclease restriction sites at which the vector may be cut and into which a desired nucleic acid sequence may be inserted. Vectors may contain one or more marker sequences suitable for use in the identification and/or selection of cells which
have or have not been transformed or genomically modified with the vector. Markers include, for example, genes encoding proteins which increase or decrease either resistance or sensitivity to antibiotics (e.g., kanamycin, ampicillin) or other compounds, genes which encode enzymes whose activities are detectable by standard assays known in the art (e.g., β-galactosidase, alkaline phosphatase or luciferase), and genes which visibly affect the phenotype of transformed or transfected cells, hosts, colonies, or plaques. Any vector suitable for the transformation of a host cell, (e.g., *E. coli*, mammalian cells such as CHO cell, insect cells, etc.) as embraced by the present invention, for example vectors belonging to the pUC series, pGEM series, pET series, pBAD series, pTET series, or pGEX series. In some embodiments, the vector is suitable for transforming a host cell for recombinant protein production. Methods for selecting and engineering vectors and host cells for expressing gRNAs and/or proteins (e.g., those provided herein), transforming cells, and expressing/purifying recombinant proteins are well known in the art, and are provided by, for example, Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)).

**DETAILED DESCRIPTION OF CERTAIN EMBODIMENTS OF THE INVENTION**

Site-specific enzymes which catalyze nucleic acid modifications are powerful tools for targeted genome modification *in vitro* and *in vivo*. Some site-specific enzymes can theoretically achieve a level of specificity for a target site that would allow one to target a single unique site in a genome for modification without affecting any other genomic site. In the case of site-specific nucleases, it has been reported that nuclease cleavage in living cells triggers a DNA repair mechanism that frequently results in a modification of the cleaved and repaired genomic sequence, for example, via homologous recombination or non-homologous end-joining. Accordingly, the targeted cleavage of a specific unique sequence within a genome opens up new avenues for gene targeting and gene modification in living cells, including cells that are hard to manipulate with conventional gene targeting methods, such as many human somatic cells or embryonic stem cells. Nuclease-mediated modification of disease-related sequences, *e.g.*, the CCR-5 allele in HIV/AIDS patients, or of genes necessary for tumor neovascularization, can be used in the clinical context, and two site-specific nucleases are currently in clinical trials (Perez, E.E. *et al.*, "Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases." *Nature Biotechnology*. 26, 808-816 (2008); ClinicalTrials.gov identifiers: NCT00842634, NCT01044654, NCT01252641, NCT01082926). Other diseases that can be treated using
site-specific nucleases or other site-specific DNA modifying enzymes include, for example, diseases associated with triplet expansion (e.g., Huntington's disease, myotonic dystrophy, spinocerebellar ataxias, etc.), cystic fibrosis (by targeting the CFTR gene), cancer, autoimmune diseases, and viral infections.

[0070] One important problem with site-specific modification is off-target effects, e.g., the modification of genomic sequences that differ from the intended target sequence by one or more nucleotides. Undesired side effects of off-target modification range from insertion into unwanted loci during a gene targeting event to severe complications in a clinical scenario. For example, off-target modification of sequences encoding essential gene functions or tumor suppressor genes may result in disease or even the death of a subject. Accordingly, it is desirable to employ new strategies in designing site-specific enzymes having the greatest chance of minimizing off-target effects.

[0071] The systems, methods, and compositions of the present disclosure represent, in some aspects, an improvement over previous methods and compositions by providing means to control the spatiotemporal activity of site-specific enzymes, for example, RNA-guided nucleases and engineered RNA-guided nucleic acid modifying enzymes. For example, RNA-guided nucleases known in the art, both naturally occurring and those engineered, typically bind to and cleave DNA upon forming a complex with an RNA (e.g., a gRNA) that complements the target. Aspects of the present invention relate to the recognition that having spatiotemporal control of the enzymatic or nucleic acid binding properties of an RNA-guided nuclease and RNA-guided nucleic acid modifying enzymes by engineering variants to include an intein will decrease the likelihood of off-target effects by minimizing or controlling the time a RNA-guided nuclease or engineered RNA-guided nucleic acid modifying enzymes is active. Accordingly, the strategies, methods, compositions, kits, and systems provided herein can be used to control the activity of any site-specific enzyme (both naturally occurring and those engineered) such as RNA-guided nucleases (e.g., Cas9, Cas9 variants, fragments of Cas9 or Cas9 variants, etc.) or engineered nucleic acid modifying enzymes comprising a variant of an RNA-guided nuclease (e.g., dCas9).

[0072] Inteins are protein splicing elements that are able to catalyze their excision out of a single polypeptide and leave behind the flanking sequences, or exteins, precisely ligated together through a native peptide bond. Inteins are attractive tools for modulating protein structure and function because they do not require any other cellular components, are able to splice out of a wide variety of extein contexts, and can undergo splicing in minutes. Although natural inteins splice spontaneously, inteins that undergo splicing in a small
molecule-dependent or ligand-dependent manner have been developed by fusing intein halves with proteins that dimerize in the presence of a small molecule, or by directed evolution in which a library of intact inteins fused to a ligand-binding domain was screened to splice in the presence, but not the absence, of a small molecule or ligand. These ligand-dependent inteins have enabled protein function in cells to be controlled post-translationally by the addition of an exogenous, cell-permeable molecule (See e.g., published U.S. Patent Application US 2014/0065711 Al, the entire contents of which are hereby incorporated by reference). The inventors have found that the targeted insertion of ligand-dependent inteins into site-specific enzymes renders the enzymes, in some instances, inactive prior to the controlled excision of the intein through binding of a ligand specific for the intein. For example, the targeted insertion of a ligand-dependent intein into Cas9 at fifteen different positions resulted in a subset of Cas9 variants that were inactive in the absence of ligand, but upon addition of the ligand the intein self-excised resulting in an active Cas9 protein capable of site-specific cleavage of a target gene.

[0073] Some aspects of this disclosure are based on the surprising discovery that Cas9 proteins comprising an intein, for example, a ligand-dependent intein as described herein, exhibit an increased specificity as compared to constitutively active Cas9 proteins. For example, it was found that the conditionally active Cas9 proteins comprising an intein exhibit an activity in the "on" state that is comparable to wild-type Cas9 activity or only slightly decreased as compared to wild-type Cas9 activity, while exhibiting decreased off-target activity.

[0074] In addition, some aspects of this disclosure relate to the recognition that Cas9 off-target activity is at least in part related to the concentration of active Cas9 proteins, and that the off-target activity of the provided conditionally active Cas9 proteins, e.g., the provided ligand-dependent Cas9 proteins, can be modulated, e.g., further decreased, by contacting the Cas9 proteins with a minimal amount of ligand effecting the desired result, e.g., the minimal amount effecting intein excision from a Cas9 protein, or the minimal amount resulting in a desired level of Cas9 protein activity.

[0075] While of particular relevance to DNA and DNA-cleaving nucleases such as Cas9 and variants thereof, the inventive concepts, methods, compositions, strategies, kits, and systems provided herein are not limited in this respect, but can be applied to any nuclease or nucleic acid:enzyme system utilizing nucleic acid templates such as RNA to direct binding to a target nucleic acid. For example, the inventive concepts provided herein can be applied to RNA-guided nucleic acid-targeting protein, e.g., to RNA-guided nucleases, and to fusion
proteins comprising nucleic acid-targeting domains of such nucleases, e.g., to fusion proteins comprising a Cas9 targeting domain (e.g., dCas9 domain), and a functional (effector) domain, such as, for example, a heterologous nuclease domain, recombinase domain, or other nucleic acid-editing domain.

**Small molecule controlled site-specific enzymes**

[0076] Some aspects of this disclosure provide site-specific enzymes engineered to have both an "on" and "off" state which depends on the presence of a ligand such as a small molecule. The ligand binds and activates the enzyme through binding a ligand-dependent intein in the enzyme, whereby ligand binding induces self-excision of the intein thereby activating the enzyme (e.g., the presence of the intein in the enzyme disrupted one or more activities of the enzyme). In some aspects then, the enzymes may collectively be referred to as "small molecule controlled" or "ligand-dependent" site-specific enzymes. In some embodiments, the site-specific enzyme that has been modified to include a ligand-dependent intein comprises Cas9, or a variant of Cas9.

[0077] Accordingly, in the absence of a ligand that binds the intein, the intein is not excised, and the protein comprising Cas9 or variant of Cas9 remains inactive. By "inactive" it is meant that the protein has no or minimal activity with respect to one or more activities described herein. In some embodiments, prior to intein excision, the protein has (i) no or minimal enzymatic activity; (ii) no or minimal gRNA binding activity; (iii) no or minimal target nucleic acid binding activity; or any combination of (i)-(iii), e.g., the protein has (i) and (ii); (i) and (iii); (ii) and (iii); or (i), (ii) and (iii). Enzymatic activities for (i), include, for example, nuclease activity, nickase activity, recombinase activity, nucleic acid editing (e.g., deaminase) activity, transcriptional activation, transcriptional repression, and epigenetic modification activity.

[0078] In some embodiments, by "minimal" activity, it is meant that the protein, prior to excision of the intein, exhibits less than 50%, less than 45%, less than 40%, less than 35%, less than 30%, less than 25%, less than 24%, less than 23%, less than 22%, less than 21%, less than 20, less than 19%, less than 18%, less than 17%, less than 16%, less than 15%, less than 14%, less than 13%, less than 12%, less than 11%, less than 10%, less than 9%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, or less than 1% of a particular activity (e.g., nuclease activity, nickase activity, recombinase activity, deaminase activity, transcriptional activation, transcriptional repression, epigenetic modification activity, gRNA binding activity, and/or target nucleic acid binding activity) as
compared to either the wild type counterpart of the protein or the intein-excised form of the protein. In some embodiments, following excision of the intein, the protein exhibits at least a 1.25-fold increase, at least a 1.5-fold increase, at least 1.75-fold increase, at least a 2.0-fold increase, at least a 2.25-fold increase, at least a 2.5-fold increase, at least a 2.75-fold increase, at least a 3.0-fold increase, at least a 3.25-fold increase, at least a 3.5-fold increase, at least a 3.75-fold increase, at least a 4.0-fold increase, at least a 4.5-fold increase, at least a 5.0-fold increase, at least a 5.5-fold increase, at least a 6.0-fold increase, at least a 6.5-fold increase, at least a 7.0-fold increase, at least a 7.5-fold increase, at least a 8.0-fold increase, at least a 8.5-fold increase, at least a 9.0-fold increase, at least a 9.5-fold increase, or at least a 10.0-fold or more increase in activity (e.g., nuclease activity, nickase activity, recombinase activity, or deaminase activity) as compared to the intein-intact form of the protein. Methods for assessing the activity of any ligand-dependent site-specific Cas9-containing enzyme provided herein are well known to those of ordinary skill in the art, and in the context of nuclease activity include those described in the Examples.

In some embodiments, upon excision, the intein leaves a cysteine residue. Thus, if the intein is inserted such that it replaces a cysteine, the Cas9 protein, upon intein excision, will be unmodified as compared to the original protein. If the intein replaces any other amino acid, the Cas9 protein, upon intein excision, will contain a cysteine in place of the amino acid that was replaced. In some embodiments, the intein does not replace an amino acid residue in a Cas9 protein, but is inserted into the Cas9 protein (e.g., in addition to the amino acid residues of the Cas9 protein). In this aspect, upon excision, the protein will comprise an additional cysteine residue. While the presence of an additional cysteine residue (or the substitution of a residue for a cysteine upon excision) is unlikely to affect the function of the Cas9 protein, in some embodiments where the intein does not replace a cysteine, the intein replaces an alanine, serine, or threonine amino acid, as these residues are similar in size and/or polarity to cysteine.

Accordingly, in some embodiments, the intein is inserted into one or both of the nuclease domains of Cas9 or a Cas9 variant (e.g., dCas9, Cas9 nickase), such as the HNH domain and/or the RuvC domain. In some embodiments, the intein is inserted into one or more other domains of Cas9 or a Cas9 variant (e.g., dCas9, Cas9 nickase), such as, REC1, REC2, PAM-interacting (PI), and/or bridge helix (BH) domain. The sequences and structure corresponding to these domains are known, and in some aspects are represented by the underlined segments of SEQ ID NO:2 (Cas9) and SEQ ID NO:5 (dCas9) above (See also, Nishimasu et al., "Crystal structure of Cas9 in complex with guide RNA and target DNA."
Cell. 2014; 156(5), 935-949). In some embodiments, the intein is inserted into any location of Cas9, e.g., any location that disrupts one or more activities of Cas9 (e.g., enzymatic activity, gRNA binding activity, and/or DNA binding activity). In some embodiments, the intein is inserted into any or more amino acids of the protein. In some embodiments, the intein replaces any cysteine, any alanine, any threonine, or any serine residue in Cas9 or a Cas9 variant including Cas9 nickase and dCas9 (and fusions thereof). In some embodiments the inserted intein sequence replaces Cys80, Ala27, Thr146, Ser219, Thr333, Thr519, Cys754, Thr622, Ser701, Ala728, Thr995, Ser1006, Ser1154, Ser1159, or Ser1274 in Cas9 (SEQ ID NO:2), dCas9 (SEQ ID NO:5), or Cas9 nickase (SEQ ID NO:4). In some embodiments, the intein is inserted within 5, within 10, within 15, or within 20 amino acids of Cys80, Ala27, Thr146, Ser219, Thr333, Thr519, Cys754, Thr622, Ser701, Ala728, Thr995, Ser1006, Ser1154, Ser1159, or Ser1274 in Cas9 (SEQ ID NO:2), dCas9 (SEQ ID NO:5), or Cas9 nickase (SEQ ID NO:4). In some embodiments, the intein sequence replaces Cys80, Ala27, Thr146, Ser219, Thr519, or Cys754 in Cas9 (SEQ ID NO:2), dCas9 (SEQ ID NO:5), or Cas9 nickase (SEQ ID NO:4). In some embodiments, a Cas9 protein comprising an intein comprises an amino acid sequence selected from the group consisting of SEQ ID NOs:27-41, or comprises an amino acid sequence that has at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% sequence identity to any one of SEQ ID NOs:27-41. In some embodiments, the intein is inserted into the protein such that it does not replace any amino acid, but is added in addition to the amino acids of the protein. The intein that is inserted into the protein can be any ligand-dependent intein, e.g., those described herein. For example, in some embodiments, the intein that is inserted into the protein comprises, in part or in whole, a sequence that is at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% identical to any one of SEQ ID NO:27-41.

Cas9: Intein (37R3-2; in double underline) replacing Cys80

MDKKSILGIDGTNSVGVAVITDEYKPSKFKVLGNTDRSHIKKNLIGALLFDSGETAEATRLKRTA
RRRYTTRKNRClAEAGTRIDFPVTGTHRIEDVDVGRKPHWAAAKDGTLLARPVSFWFDGTRVDI
GLRIAGGAIVWATPDHKVLTEYGWRAAGELRKGDVAGPGSGSNLSLTDADQVMVSAALDDAPFILY
SEYDPTSFSEASMGLLTNLADRELVHMINWAKRVPFGVDLHDAQHALERAWLEILMIGLVWRSME
EHFGKLLFAPNLLDLRDNQGKCVEGMEIFDMLLATSSRFRMMLQGEEVFVLKSIILLNSGYTFLLSS
TLKSLLEKDHTRALDKTDTLIHMAMKAGTLQQHQORLQLIILSLHJRMDNSKGMEMHLYSMKYKN
WPLYDDLMILDARLHAGGSGASRVRQAFADALDDDDKFLHDLAEGLRSVIREVLPTRRARTFDLEG
EELHTLVAEGVWHNCY_IQEIFSNEKAVDDSFHRLEESFLVEEDKKHERHFIGNIVDEVAYHEKY
PTIYHLKLVSTDKADRLLYSLALAHMIKFRGHLFIEGDLNFDNVDKVFLIQVLQYNEFVPEN
INASGVDKAILLSLQSRRENIQLPFHKKNGLFLGLIALSLGTLPFNSFDLAEADKLQLQSK
DTYDDDLDNLQIQDQYADLFLAANLSDAILLSDILDRLVNEITKAPLSASMKRYDEHQQDLTLK
ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT
FDNSIPHFQILHELHAIRLRRQEDFYPFLKDNREJEIEITFRIPYYVGPRLGNRsRAWMTKSEET
ITPWNFEEWDKGASAGSFIFERMNTNDFKMLEVKPHSHLYEYTFVNNELTVKVVTEGMRKPFAFL
SQQKKAIVDLDFFKTNRKVTQKVLKQEDFKKEIEFSDFSVEI SGVEDRFNASLGTHDLKLI IKDKKDFLD
NEEEDILEDIVLITLIFEDREMIEERLKYAHFLDDKVKMLKRRYTGWRLSRKLINGIRDKQSG
KTILDLFSKDFSNFKNQLHDDSLTFIKEDFQIAQVSVQQGSDLHEHIANLSPAIKGLQTKVKV
DELVKVMGRHPENIVIERAMENQGTTQKGNRSNRMKRIIEEGKELGSQLKLEHVENTQLNQELYK
LYLQLQGNYRMDVNLRLSDYDDVHISPSFLKSDKSINDVLRKTSNDRKGSNVDSPSEEEVKKMK
NYMRQQLNKLITQRKFDNLTKEAGRLSDKALDFIKRFQKLQVELTRQITKHAQIILDSMTNKYENDK
LIREVKVTLKSLVSDFRKDQFYFVKEVINXYHHHADYALNAVGTALIIKYPKLESEFVGYDKVY
DVRKMIASKEQEIGKATAKYFNSMNFKTEITLANGEIRKPLIETNGETGEIVWDRKGDFATVR
KVLSPMQNVIVKKTVEQTVGFGSE ILKKNRSOKLARDDWDPPKGYGDSTAYSYLVLVAEKEG
KSKKLKSVKLELGITIMERSFSEKTFDLEAKGYKVEVKDLKI ILKPLYSKLFLENGKRKMLASAGEL
QKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFKRVLADNLKDVLSSAYNKHRDKPIREQAENI ILHFILTNLGAFAFKEYFDDTIDDSRKRTSTKVLATL
HIQS ITGLY
ETRIDLSQGDD (SEQ ID NO:27)

Cas9:Intein (37R3-2; in double underline) Alal27
MDKKSIGLIDGTNSVGAWVITDEYKVSFKKFLGNIDRRHSIKKNLIGALLFDGETEAAITLKRITA
RRYTRKRNKCLYQIEFNSMIAKVDSSFHFLSLVEFLVEEDKKHERPIFGNIVEDE CLEAEGTRIED
PVYTGTHIERWDGRKPIHVAAKGTDLARRPSWFDGTDVIGLIRIAGAIWATPDKHVLITE
YGWRAAGELRKDRVAGPPGGSNALSALITADQMSALDDEPPILYSEYDPTSFSEASMMGLLTRLN
ADRELVHINWAKRVPVFVDLTLHDAQHILLERALWELILMIGLVWRSMEHPFGKLLFAPNILDLLNNQGC
EGVMEVEIFDMALAYASSRMRNMLQGEVCVNL IILLNSGVTFLSTLKSLEEDHIHALDKTIDT
LIHLMARAGKTLQHQHRQLAQLLLSISRHMSNKMGHELYSMKYNVQPFYLLEMLDRAHRLHAGG
SGASVQAFQADLKHMGLAELEGLYRSVOURPRTLHRDFTEVEELHTVGLAVWNGVHCYK
PTIYHLRRKLVDSRDKADLLRIYIALAHMLFRRHFLIEGDLNPDNDSVDKFLIQLVTQYNLIEEEEP
INASGVDIAKILSALKSRSRLNIAQQLPGEKGNKFLGNNLIALGLTPLNKSFNFDLDNADKGQSLK
DITYDDDDNLALLAQIGQDYADILFAAKLNSAIDLSDILVRNTIEITKAPLASMIFKRYDEHQDLTLK
ALVRQQLPKEIFDFQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT
FDNSIPHFQILHELHAIRLRRQEDFYPFLKDNREJEIEITFRIPYYVGPRLGNRsRAWMTKSEET
ITPWNFEEWDKGASAGSFIFERMNTNDFKMLEVKPHSHLYEYTFVNNELTVKVVTEGMRKPFAFL
SQQKKAIVDLDFFKTNRKVTQKVLKQEDFKKEIEFSDFSVEI SGVEDRFNASLGTHDLKLI IKDKKDFLD
NEEEDILEDIVLITLIFEDREMIEERLKYAHFLDDKVKMLKRRYTGWRLSRKLINGIRDKQSG
KTILDLFSKDFSNFKNQLHDDSLTFIKEDFQIAQVSVQQGSDLHEHIANLSPAIKGLQTKVKV
DELVKVMGRHPENIVIERAMENQGTTQKGNRSNRMKRIIEEGKELGSQLKLEHVENTQLNQELYK
LYLQLQGNYRMDVNLRLSDYDDVHISPSFLKSDKSINDVLRKTSNDRKGSNVDSPSEEEVKKMK
NYMRQQLNKLITQRKFDNLTKEAGRLSDKALDFIKRFQKLQVELTRQITKHAQIILDSMTNKYENDK
LIREVKVTLKSLVSDFRKDQFYFVKEVINXYHHHADYALNAVGTALIIKYPKLESEFVGYDKVY
DVRKMIASKEQEIGKATAKYFNSMNFKTEITLANGEIRKPLIETNGETGEIVWDRKGDFATVR
KVLSPMQNVIVKKTVEQTVGFGSE ILKKNRSOKLARDDWDPPKGYGDSTAYSYLVLVAEKEG
KSKKLKSVKLELGITIMERSFSEKTFDLEAKGYKVEVKDLKI ILKPLYSKLFLENGKRKMLASAGEL
QKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFKRVLADNLKDVLSSAYNKHRDKPIREQAENI ILHFILTNLGAFAFKEYFDDTIDDSRKRTSTKVLATL
HIQS ITGLY
ETRIDLSQGDD (SEQ ID NO:28)

Cas9:Intein (37R3-2; in double underline) Thrl46
MDKKSIGLIDGTNSVGAWVITDEYKVSFKKFLGNIDRRHSIKKNLIGALLFDGETEAAITLKRITA
RRYTRKRNKCLYQIEFNSMIAKVDSSFHFLSLVEFLVEEDKKHERPIFGNIVEDE CLEAEGTRIED
PVYTGTHIERWDGRKPIHVAAKGTDLARRPSWFDGTDVIGLIRIAGAIWATPDKHVLITE
YGWRAAGELRKDRVAGPPGGSNALSALITADQMSALDDEPPILYSEYDPTSFSEASMMGLLTRLN
ADRELVHINWAKRVPVFVDLTLHDAQHILLERALWELILMIGLVWRSMEHPFGKLLFAPNILDLLNNQGC
EGVMEVEIFDMALAYASSRMRNMLQGEVCVNL IILLNSGVTFLSTLKSLEEDHIHALDKTIDT
LIHLMARAGKTLQHQHRQLAQLLLSISRHMSNKMGHELYSMKYNVQPFYLLEMLDRAHRLHAGG
SGASVQAFQADLKHMGLAELEGLYRSVOURPRTLHRDFTEVEELHTVGLAVWNGVHCYK
PTIYHLRRKLVDSRDKADLLRIYIALAHMLFRRHFLIEGDLNPDNDSVDKFLIQLVTQYNLIEEEEP
INASGVDIAKILSALKSRSRLNIAQQLPGEKGNKFLGNNLIALGLTPLNKSFNFDLDNADKGQSLK
DITYDDDDNLALLAQIGQDYADILFAAKLNSAIDLSDILVRNTIEITKAPLASMIFKRYDEHQDLTLK
ALVRQQLPKEIFDFQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT
FDNSIPHFQILHELHAIRLRRQEDFYPFLKDNREJEIEITFRIPYYVGPRLGNRsRAWMTKSEET
ITPWNFEEWDKGASAGSFIFERMNTNDFKMLEVKPHSHLYEYTFVNNELTVKVVTEGMRKPFAFL
SQQKKAIVDLDFFKTNRKVTQKVLKQEDFKKEIEFSDFSVEI SGVEDRFNASLGTHDLKLI IKDKKDFLD
NEEEDILEDIVLITLIFEDREMIEERLKYAHFLDDKVKMLKRRYTGWRLSRKLINGIRDKQSG
KTILDLFSKDFSNFKNQLHDDSLTFIKEDFQIAQVSVQQGSDLHEHIANLSPAIKGLQTKVKV
DELVKVMGRHPENIVIERAMENQGTTQKGNRSNRMKRIIEEGKELGSQLKLEHVENTQLNQELYK
LYLQLQGNYRMDVNLRLSDYDDVHISPSFLKSDKSINDVLRKTSNDRKGSNVDSPSEEEVKKMK
NYMRQQLNKLITQRKFDNLTKEAGRLSDKALDFIKRFQKLQVELTRQITKHAQIILDSMTNKYENDK
LIREVKVTLKSLVSDFRKDQFYFVKEVINXYHHHADYALNAVGTALIIKYPKLESEFVGYDKVY
DVRKMIASKEQEIGKATAKYFNSMNFKTEITLANGEIRKPLIETNGETGEIVWDRKGDFATVR
KVLSPMQNVIVKKTVEQTVGFGSE ILKKNRSOKLARDDWDPPKGYGDSTAYSYLVLVAEKEG
KSKKLKSVKLELGITIMERSFSEKTFDLEAKGYKVEVKDLKI ILKPLYSKLFLENGKRKMLASAGEL
QKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFKRVLADNLKDVLSSAYNKHRDKPIREQAENI ILHFILTNLGAFAFKEYFDDTIDDSRKRTSTKVLATL
HIQS ITGLY
ETRIDLSQGDD (SEQ ID NO:28)
Cas9: Intein (37R3-2; in double underline) replacing Ser219

<table>
<thead>
<tr>
<th>Cas9: Intein (37R3-2; in double underline) replacing Ser219</th>
</tr>
</thead>
</table>

Cas9: Intein (37R3-2; in double underline) replacing Ser219

<table>
<thead>
<tr>
<th>Cas9: Intein (37R3-2; in double underline) replacing Ser219</th>
</tr>
</thead>
</table>

Cas9: Intein (37R3-2; in double underline) replacing Ser219

<table>
<thead>
<tr>
<th>Cas9: Intein (37R3-2; in double underline) replacing Ser219</th>
</tr>
</thead>
</table>

Cas9: Intein (37R3-2; in double underline) replacing Ser219

<table>
<thead>
<tr>
<th>Cas9: Intein (37R3-2; in double underline) replacing Ser219</th>
</tr>
</thead>
</table>

Cas9: Intein (37R3-2; in double underline) replacing Ser219

<table>
<thead>
<tr>
<th>Cas9: Intein (37R3-2; in double underline) replacing Ser219</th>
</tr>
</thead>
</table>

Cas9: Intein (37R3-2; in double underline) replacing Ser219

<table>
<thead>
<tr>
<th>Cas9: Intein (37R3-2; in double underline) replacing Ser219</th>
</tr>
</thead>
</table>

Cas9: Intein (37R3-2; in double underline) replacing Ser219

<table>
<thead>
<tr>
<th>Cas9: Intein (37R3-2; in double underline) replacing Ser219</th>
</tr>
</thead>
</table>

Cas9: Intein (37R3-2; in double underline) replacing Ser219

<table>
<thead>
<tr>
<th>Cas9: Intein (37R3-2; in double underline) replacing Ser219</th>
</tr>
</thead>
</table>

Cas9: Intein (37R3-2; in double underline) replacing Ser219

<table>
<thead>
<tr>
<th>Cas9: Intein (37R3-2; in double underline) replacing Ser219</th>
</tr>
</thead>
</table>

Cas9: Intein (37R3-2; in double underline) replacing Ser219

<table>
<thead>
<tr>
<th>Cas9: Intein (37R3-2; in double underline) replacing Ser219</th>
</tr>
</thead>
</table>

Cas9: Intein (37R3-2; in double underline) replacing Ser219

<table>
<thead>
<tr>
<th>Cas9: Intein (37R3-2; in double underline) replacing Ser219</th>
</tr>
</thead>
</table>

Cas9: Intein (37R3-2; in double underline) replacing Ser219

<table>
<thead>
<tr>
<th>Cas9: Intein (37R3-2; in double underline) replacing Ser219</th>
</tr>
</thead>
</table>

Cas9: Intein (37R3-2; in double underline) replacing Ser219

<table>
<thead>
<tr>
<th>Cas9: Intein (37R3-2; in double underline) replacing Ser219</th>
</tr>
</thead>
</table>

Cas9: Intein (37R3-2; in double underline) replacing Ser219

<table>
<thead>
<tr>
<th>Cas9: Intein (37R3-2; in double underline) replacing Ser219</th>
</tr>
</thead>
</table>

Cas9: Intein (37R3-2; in double underline) replacing Ser219

<table>
<thead>
<tr>
<th>Cas9: Intein (37R3-2; in double underline) replacing Ser219</th>
</tr>
</thead>
</table>
Cas9: Intein (37R3-2; in double underline) replacing Thr333

MDKYS IGLDIGNSVGWAVITDEYKPSKKFVGLNTRDSISIKNLIGALLFSDSGETAATELRKRTA
RRYTRKRNCIRYCLEIFSNEFNSAVDDFDHHLFRRLSLVEEDKHKHERPHFNGIDVAYHEKTYI
HLRLKVLSDTKDAKLRLI YLALAHIKFCRHFLIEGLDNPSDVKLFIQLVQTYNLQFNEPINAS
GVDKAKILLSKLRRLELQAQLPGEQKKNGFLNLALLSLILTPNFKNSDFLAE DAKLQLSKTYD
DDLLNLAAQGIDYADLFIAAKNLSDAILLSDILRNTVEITKAPLASMIRKDYEHQDL CIAEGRF
FPVTGTHIRGVEDPFKPGSNAALSAITDAQVMSALDAAEPILYSEITPFSFSEAMGMLLT
NLADRELNVINMAWAKRPVGPFDVLLHDQHAHLERWLEIIMLGWRSMHGPQLFLAP.currentTimeMillis
KCEVGMEVIFDMLATTSSRMNNLQEEGFCVLCISKI ILLNSGVTFLSLEKSKEDHIIHRALDICT
DTLHILMAKAGLTLQQHQRLQLLLLILSRHRMSKGMELSYMKYNWYDPDDLMLEMAHRHLA
GGGSARQVFAFADALDKFLHDMALAGRRYSVIRELPRTTARFDLVEELHHTLVEAGWVVHNCLL
ALVRQLEPKYKEIIFDSQSKNGAYIGDDGASQQEAYFYKIPKILEMGTDEELVKNREDILLKRQRT
FDNQSHIQHILHGLEHAIHRQDDEPYFLKNDRKEIEKILTRFIPYVGALRNSFWMRKTSSEET
ITPFWNEEWDKAGSAQSFIERTMTDKNLPVNKLPEKPHSLLEYITYFTYNLEITKVKYTVEMGRKPAFL
SQEQQKAIYDLPFKTRKVTQVKEDKQFIEFCDSVEI SVEFDRFNASLGYTHDLIK IKDFFDL
NEENILEDLVIDILITLFEDREMEERLTYAHLFDVVMKQLKRRTYTGWLRSLKINGIRDSQG
KTILDLFDSSGMDFHSLDLITDEKQIAVQGSQDGSLHEAHIAINSPAIGKILGTQKV
DELVKVRGHKPIVENIMERANETQTKQKGNSRMRKKEIEEGKELSGILKHEFVIENTLQNKELY
LYLQNGRMDVQDELDINRLSDYDVHIVPQSFKDDSIDNKVLTRSDKNGKDSDNSPVEEIVKKMK
NYQMRVLNNAKLTTQKFDNFTKAEKFLDAKQIFKQLVETRQITKHVAILLDSRMNTKYEENDK
LIREVKVLTILKVLSDFRDFQFKYKREINNYNDAYADDNVATGALIYKPKLESEFYVGVKY
DVRMIAKSEQKEREITLANGEIRKPLIEGNETEVGWDGREAPTV
KVLSPMQVNVKKEVTQGTFCGSFES ILPKRNSDKLIARKWDPPKYGGFDSPTVAYSVLVVAKEVCG
KKKLKSVKELLGITIMERSSEFENPFDIFLEAKYGKEVKKDLI ILPKYSLEENKRMLASAGEL
QKGNELALFSKYVLYLAHYSJEKLSGKPDNEQQLFVQHKKYLYDEI IEQISEFSKVILADANLD
KVSALNKRDKPISREQAEIHLFLTNLGAFAFPKYFTIDRTKRTYSTKEDVIALTHQ ITGELY
ETRIDLSQLGDD (SEQ ID NO:31)

Cas9: Intein (37R3-2; in double underline) replacing Thr519

MDKYS IGLDIGNSVGWAVITDEYKPSKKFVGLNTRDSISIKNLIGALLFSDSGETAATELRKRTA
RRYTRKRNCIRYCLEIFSNEFNSAVDDFDHHLFRRLSLVEEDKHKHERPHFNGIDVAYHEKTYI
HLRLKVLSDTKDAKLRLI YLALAHIKFCRHFLIEGLDNPSDVKLFIQLVQTYNLQFNEPINAS
GVDKAKILLSKLRRLELQAQLPGEQKKNGFLNLALLSLILTPNFKNSDFLAE DAKLQLSKTYD
DDLLNLAAQGIDYADLFIAAKNLSDAILLSDILRNTVEITKAPLASMIRKDYEHQDL CIAEGRF
FPVTGTHIRGVEDPFKPGSNAALSAITDAQVMSALDAAEPILYSEITPFSFSEAMGMLLT
NLADRELNVINMAWAKRPVGPFDVLLHDQHAHLERWLEIIMLGWRSMHGPQLFLAP.currentTimeMillis
KCEVGMEVIFDMLATTSSRMNNLQEEGFCVLCISKI ILLNSGVTFLSLEKSKEDHIIHRALDICT
DTLHILMAKAGLTLQQHQRLQLLLLILSRHRMSKGMELSYMKYNWYDPDDLMLEMAHRHLA
GGGSARQVFAFADALDKFLHDMALAGRRYSVIRELPRTTARFDLVEELHHTLVEAGWVVHNCLL
ALVRQLEPKYKEIIFDSQSKNGAYIGDDGASQQEAYFYKIPKILEMGTDEELVKNREDILLKRQRT
FDNQSHIQHILHGLEHAIHRQDDEPYFLKNDRKEIEKILTRFIPYVGALRNSFWMRKTSSEET
ITPFWNEEWDKAGSAQSFIERTMTDKNLPVNKLPEKPHSLLEYITYFTYNLEITKVKYTVEMGRKPAFL
SQEQQKAIYDLPFKTRKVTQVKEDKQFIEFCDSVEI SVEFDRFNASLGYTHDLIK IKDFFDL
NEENILEDLVIDILITLFEDREMEERLTYAHLFDVVMKQLKRRTYTGWLRSLKINGIRDSQG
KTILDLFDSSGMDFHSLDLITDEKQIAVQGSQDGSLHEAHIAINSPAIGKILGTQKV
DELVKVRGHKPIVENIMERANETQTKQKGNSRMRKKEIEEGKELSGILKHEFVIENTLQNKELY
LYLQNGRMDVQDELDINRLSDYDVHIVPQSFKDDSIDNKVLTRSDKNGKDSDNSPVEEIVKKMK
NYQMRVLNNAKLTTQKFDNFTKAEKFLDAKQIFKQLVETRQITKHVAILLDSRMNTKYEENDK
LIREVKVLTILKVLSDFRDFQFKYKREINNYNDAYADDNVATGALIYKPKLESEFYVGVKY
DVRMIAKSEQKEREITLANGEIRKPLIEGNETEVGWDGREAPTV
KVLSPMQVNVKKEVTQGTFCGSFES ILPKRNSDKLIARKWDPPKYGGFDSPTVAYSVLVVAKEVCG
KKKLKSVKELLGITIMERSSEFENPFDIFLEAKYGKEVKKDLI ILPKYSLEENKRMLASAGEL
Cas9: Intein (37R3-2, in double underline) replacing Cys574

MDKKYS IGLDIGNSTGVAWITDEYKVPSKFKVFVLGNTDRHISKKNLIGALLFDSGETAEATRLKRTA
RYYYTRKRKNRCYLCQIFSNEKAVDSDHFRHELLESFLVEEDKKHERHPIFGNIDEVAYHEKYTIY
HLRKVLDFSSTLDRLRLY ALALAHMIFGHFLEGILNPNDSDVFKLDLFOVLQTVTNLQFEEPNINAS
GVDKAIALSRLSKSRLENLAQILPQGEEKMNLGIALGGLTNFSKNDLAE DAKLQSKDTYD
DSDLNLAIGQDQALDFLALKNLAILSDDLLSLSDILLRVNETIKAPLSAIMEKRYDEHQDLLLKLALKVR
QLQPEKYEIFDFQSKNGYAGYIDGGASQEEFYKFKPEQKIQPIKMDGTEELLVNLRLKKLRQFTDNG
SIPOHQLGHLEILHRNRQEDFQFPLKDFNKREILKTLFYPVPLGRSRAMNFKMTSSEEVTVKMK
NMYRQNLLNLKQKLDLTTAKRNFDRHLCFALQVDSKQFELTDKQFEDQMDRNLKLRQ KorHDKQSKG
LIREKVITLSSKVLQDFQYFQKREINNYHHADHAYLNAVGTALKYPKLESEFVYGKVLVR
DVRMMKQAEQIEQIGATKAYFFYSNEKMFKETILEIANGEIRKRPLEINTEIGVIVKGRDFATVR
KVLSSPQVNIVKTTQGFSGKES ILPKRNSDQLLARKWDSDKPGFFDSTPVAYSVLVVAKVEKG
KSKLKSVKELLGITTIMSFFEKNPIDLEAKVEKVKDLI IMLPKSYLFELENGRKMLASAGEL
QKGNLALSFKVNYLMASHYELKLGSPEDNQKQLFVEQKHYLDIEI EQISEFSKRVILADANLD
KVLSAYNKHRDKPIREQAENI IHLFTLNLGAFAAFKYFDTTIDRKRKYTSTKEVLDAI LHSQ ITGLY
ETRIDLQSLQGD (SEQ ID NO:32)

Cas9: Intein (37R3-2, in double underline) replacing Thr622

MDKKYS IGLDIGNSTGVAWITDEYKVPSKFKVFVLGNTDRHISKKNLIGALLFDSGETAEATRLKRTA
RYYYTRKRKNRCYLCQIFSNEKAVDSDHFRHELLESFLVEEDKKHERHPIFGNIDEVAYHEKYTIY
HLRKVLDFSSTLDRLRLY ALALAHMIFGHFLEGILNPNDSDVFKLDLFOVLQTVTNLQFEEPNINAS
GVDKAIALSRLSKSRLENLAQILPQGEEKMNLGIALGGLTNFSKNDLAE DAKLQSKDTYD
DSDLNLAIGQDQALDFLALKNLAILSDDLLSLSDILLRVNETIKAPLSAIMEKRYDEHQDLLLKLALKVR
QLQPEKYEIFDFQSKNGYAGYIDGGASQEEFYKFKPEQKIQPIKMDGTEELLVNLRLKKLRQFTDNG
SIPOHQLGHLEILHRNRQEDFQFPLKDFNKREILKTLFYPVPLGRSRAMNFKMTSSEEVTVKMK
NMYRQNLLNLKQKLDLTTAKRNFDRHLCFALQVDSKQFELTDKQFEDQMDRNLKLRQ KorHDKQSKG
LIREKVITLSSKVLQDFQYFQKREINNYHHADHAYLNAVGTALKYPKLESEFVYGKVLVR
DVRMMKQAEQIEQIGATKAYFFYSNEKMFKETILEIANGEIRKRPLEINTEIGVIVKGRDFATVR
KVLSSPQVNIVKTTQGFSGKES ILPKRNSDQLLARKWDSDKPGFFDSTPVAYSVLVVAKVEKG
KSKLKSVKELLGITTIMSFFEKNPIDLEAKVEKVKDLI IMLPKSYLFELENGRKMLASAGEL
QKGNLALSFKVNYLMASHYELKLGSPEDNQKQLFVEQKHYLDIEI EQISEFSKRVILADANLD
KVLSAYNKHRDKPIREQAENI IHLFTLNLGAFAAFKYFDTTIDRKRKYTSTKEVLDAI LHSQ ITGLY
ETRIDLQSLQGD (SEQ ID NO:33)

Cas9: Intein (37R3-2, in double underline) replacing Thr622
Cas9: Intein (37R3-2; in double underline) replacing Ser701
MDKKYSIGLDIGTNSGVWAVITDEYKPSKKFKVLGNTDHRSIKNNLIGALLFDSGETAEATRLKRTA
RRRTKRNKRICYLFIEFSNEMAKVDSFFHRLEESFLVEEDKKHERPHFGNIVDEVAYHEKYPTIT
HLRRKVLVDSTDKADLRIIYLALHMIFKRGHFLIEGDLDNPNSVDKLFIQLVTYQNLFEENPINAS
GVDAKALSARLKSRRLENLIQAQLPGEKKNLGFNLIALSGLTNPFSNFLDLAE DAKLQLKSTDYT
DDDLNLAAIQGQADIFLAAKNLSAALLSLDILVRNTEITKAPLASSMIRKRYEHDQDLTLLKALVR
QQLPEKKYIEFFDPQSGKNGAYIGIDGSAEQEFYKFIPKILEKMDGTEELVKLNRDLDRKRTTFDNG
SIPHQIFHLGELHAIHLRQEDIFLKNREIEKILTFRIFYVGPRLNRSFWMTRKSETETTFW
NFEENMIKAGSAQSFIEKQMNFKPVLNPKHSLLYEYFTPVNELTVEKVKYVTGEMRKPAPFLSGEQ
KAIADVLLFLKTRVKTQVLKQKEMKFFCEFDSEISVGEDFNASLGTYHLDDL IKDKFLDNENEE
EDILEDIVITFMEMIEDRELKERYSTFNYDHVKKSLRKLINGIDQSGKITL
DFLKSDFANRMQQLHDDCLAEITGRFDVTSEITHTEDWDRIISDIKAAKDTLILARFWSW
FDQGTRDVIGLRIIAGIAVIAWTPDHKLXYEGLYGRAEGELRKGDRVAPGGSNSLALSILTDQMVSAL
LDAEFFPILFYEDPPTSFEASMGLGLNLNLADVLELVHINMMWKARPVGFVGLLTHDAHLLARWEIL
MAGLWRSMEHPGKLLFIPALLLNRDNQGCVEGMEIFDMLATSSRFRMNQLGEFVCLKS IILN
SGVTFLSIOLSLKKEEDHIIHALKTDITTLIIHIAMAGLTLQQOHRLLAQLLLILLHIIRHASNKME
HLHMSKYYVRVPYDLLELMELDLHAGGSGASVRQAFADLLDDKFLHDMLAEGLRYSEVIRELPTR
RATFDFLEVELHTLVAEGVWHNC LTFKEDIQAQKVQSGDSLHELTHANLAGPSIAKIGILQTVKVV
DELVKVMGHRPKENIVIEMARENQTQKQGKSNRMRKIEEGK ElleQGSLQKLEHVENTQLNUKELY
LLYLNQGRNMDYQDELDINRLSDDYDHIQPVSFLKDSSDNKVRKDSNVPSEEEVKMK
NYWRQQLNKLITQRFDFNLKAERGLSELSDKAGFKIRQLVEIRQITKHVAQIDSERNPTYKENDK
LIREVKVITLKSVLSDFRDFQFQYKREINIHHAYHADAYLNAVGTALIKYFKLESEFVYGYDKVY
DVRMIKAQSEIQIGKATAKYFFYESMNFKEYTEITLANGEIRKRPLIETINGEGIEIVWDKRGDFATVR
KVLSMPQVNIVKTEQVQGFSESL IPRLRSNLKLARLKDDWPDPKYGGFDSTPAYSVVLVAKVEK
KSKLKSVEKLLGDTMERSFSFKNFIDLEAGKYEVKVDIKL DKPLYSLFLENGRKRMASAGE
LQKNELALPSYFNLYASHYEKLKSGPEDNEQKQLFQVEQHKYLYDE IEQISeFSRKLADANLD
KVLSAYNHKRPDKREGAEN1 HLHTLNLGAFAAFKVDTIDDRTKRYSTKEVLATL IHQS ITGLY
ETRIDLSSQLGDD (SEQ ID NO:34)

Cas9: Intein (37R3-2; in double underline) replacing Ala728
MDKKYSIGLDIGTNSGVWAVITDEYKPSKKFKVLGNTDHRSIKNNLIGALLFDSGETAEATRLKRTA
RRRTKRNKRICYLFIEFSNEMAKVDSFFHRLEESFLVEEDKKHERPHFGNIVDEVAYHEKYPTIT
HLRRKVLVDSTDKADLRIIYLALHMIFKRGHFLIEGDLDNPNSVDKLFIQLVTYQNLFEENPINAS
GVDAKALSARLKSRRLENLIQAQLPGEKKNLGFNLIALSGLTNPFSNFLDLAE DAKLQLKSTDYT
DDDLNLAAIQGQADIFLAAKNLSAALLSLDILVRNTEITKAPLASSMIRKRYEHDQDLTLLKALVR
QQLPEKKYIEFFDPQSGKNGAYIGIDGSAEQEFYKFIPKILEKMDGTEELVKLNRDLDRKRTTFDNG
SIPHQIFHLGELHAIHLRQEDIFLKNREIEKILTFRIFYVGPRLNRSFWMTRKSETETTFW
NFEENMIKAGSAQSFIEKQMNFKPVLNPKHSLLYEYFTPVNELTVEKVKYVTGEMRKPAPFLSGEQ
KAIADVLLFLKTRVKTQVLKQKEMKFFCEFDSEISVGEDFNASLGTYHLDDL IKDKFLDNENEE
EDILEDIVITFMEMIEDRELKERYSTFNYDHVKKSLRKLINGIDQSGKITL
DFLKSDFANRMQQLHDDCLAEITGRFDVTSEITHTEDWDRIISDIKAAKDTLILARFWSW
FDQGTRDVIGLRIIAGIAVIAWTPDHKLXYEGLYGRAEGELRKGDRVAPGGSNSLALSILTDQMVSAL
LDAEFFPILFVYEDPPTSFEASMGLGLNLNLADVLELVHINMMWKARPVGFVGLLTHDAHLLARWEIL
MAGLWRSMEHPGKLLFIPALLLNRDNQGCVEGMEIFDMLATSSRFRMNQLGEFVCLKS IILN
SGVTFLSIOLSLKKEEDHIIHALKTDITTLIIHIAMAGLTLQQOHRLLAQLLLILLHIIRHASNKME
HLHMSKYYVRVPYDLLELMELDLHAGGSGASVRQAFADLLDDKFLHDMLAEGLRYSEVIRELPTR
RATFDFLEVELHTLVAEGVWHNC LTFKEDIQAQKVQSGDSLHELTHANLAGPSIAKIGILQTVKVV
DELVKVMGHRPKENIVIEMARENQTQKQGKSNRMRKIEEGK ElleQGSLQKLEHVENTQLNUKELY
LLYLNQGRNMDYQDELDINRLSDDYDHIQPVSFLKDSSDNKVRKDSNVPSEEEVKMK
NYWRQQLNKLITQRFDFNLKAERGLSELSDKAGFKIRQLVEIRQITKHVAQIDSERNPTYKENDK
LIREVKVITLKSVLSDFRDFQFQYKREINIHHAYHADAYLNAVGTALIKYFKLESEFVYGYDKVY
DVRMIKAQSEIQIGKATAKYFFYESMNFKEYTEITLANGEIRKRPLIETINGEGIEIVWDKRGDFATVR
KVLSMPQVNIVKTEQVQGFSESL IPRLRSNLKLARLKDDWPDPKYGGFDSTPAYSVVLVAKVEK
KSKLKSVEKLLGDTMERSFSFKNFIDLEAGKYEVKVDIKL DKPLYSLFLENGRKRMASAGE
LQKNELALPSYFNLYASHYEKLKSGPEDNEQKQLFQVEQHKYLYDE IEQISeFSRKLADANLD
KVLSAYNHKRPDKREGAEN1 HLHTLNLGAFAAFKVDTIDDRTKRYSTKEVLATL IHQS ITGLY
ETRIDLSSQLGDD (SEQ ID NO:35)
Cas9: Intein (37R-2; in double underline) replacing Thr995

MDKKS1GIDGNTSVGAVITIDEYKVFKKFKVGLNGTDRSHSKKNNLIGALLFDGSETAEEATLRSTA
RRYTRYRRKNRICYLQEISFNEMAKVDSSFHHLREESFLVEEKHERHPFGNIDVEAYHEKPYTIYLHLRKLVDSDKLGFRHFLIEGDLNPSVDKLFLIQLVQNFLEEPINAS
GVDAAKALSLLKSSRLENLIAQLPGKEKKNGLFLGLNALSGLTNPFSNFDLPAILDKLQSLKDTRYDDDLNLLAQIGDQYADLFAAKNLSDLILLDLIRNEITKAPLASISMRKYDEHQLITLLILKVR
QQLFPEKYEIFDFFQSKNMYAGGQTLKSYEPPFYIKFLKEMDGTVEELVLKINREDLRLRRKQTFDNG
SPIPHQIHLHAILRQRQEDYYFPLKDRNREIKELTRIFIPYYVPLARNSRFWAMTRKSEETITPF
NFEENWDGKASAQSFIERMTNFDKNPSQELKVPSLKKLLSYETFVYNETLKTVYKGMPKLSGFSEM
KAILDLFLKTNRKVTKVQKEDFYKIEFDSVEISGVEDRFNASLGTYHDLKIIKDKFDLNEEDE
DILEDIVLFLEDREMYAHLOFTDKVMQLLPFRRYTGWGRSRLKINGIRDSQKSTIIL
DFLKSDGQANRMRFSQILHDDSITFKIDQAKOSQGQLSLHEHIANLPSAUIGKIOLTVQKVDDEL
KVGRRHKPENIVIEMARENQYTTQGKQNERSMRKIEEIKELGSQILKEHPVENTQLONEKYLLYL
QNGRMDYMVDQELDNLSSDLVDHPIDNKLDSIDNKLCDTRSNKGSVDPEEVEKMKMNH
QLNNAKLLITQRFDNLTKAERGKSDLKAGFIKRQOLVETRQKTHAVQILDLSMTNKFYDENDLIERK
VKVTILSKLVSDFKDQFYKVEEINNYHAHDAYLAENVNG CLEAGTRIDFPTVGTREIDIHugh Amy
GPQGSGSLALSITADQVMSALLDAEFFPYSEYDTSFSPEASMMMGLLTLNADRELVMINAWKRPV
GFVDLTHQDLHERALWELIMIGLWRSMEPHPKLFAJPNLLDNRONGQGVEGMRVEFDMLLASS
RFRMNNQLEGEVFCISKI ILLNSGVITFLSLKLSEKEDHIHALKDTILDILHARGLTQLQHQC
QHRLAQLILLSHIRSNKSGMEHSLSKYWNPLYLIMELDAHRLHAGGASGARQAPADALDDK
FLHMLAEGLNYSVIREPLTRARDFLEVELHTLAEEVWNHCALIJKLMNOPSEFYSVGYKY
DVRVMIAKSEQIEKFKYSNIMNFETEITLANEEIRKPLEIETNGTEIWEVDKGRDFATVR
KVLSMPVQNVITEQVGFGSKSERIKRNRSKIARDDKWKPGDFSPTAVSVLWAKVEKG
KSSKLKSVKELLGITIMERSPEKNPQIDLEAKGKVEKKDLI IKLPSYLSELENGRKRMLASGAGEL
QKGKNAELPSKYVFNLYLASHEYKLSGSPDNEQKQLFVEQHKYXDEIEIQISEFPSKVILADANL
KVLASYNKRHKDPIREQAENI IHLFTLTNGAPAAFYDFTIDRKYRSTKEVLDATL IHQS ITGLY
ETRIDLSQGDL

Cas9: Intein (37R-2; in double underline) replacing Ser1006

MDKKS1GIDGNTSVGAVITIDEYKVFKKFKVGLNGTDRSHSKKNNLIGALLFDGSETAEEATLRSTA
RRYTRYRRKNRICYLQEISFNEMAKVDSSFHHLREESFLVEEKHERHPFGNIDVEAYHEKPYTIYLHLRKLVDSDKLGFRHFLIEGDLNPSVDKLFLIQLVQNFLEEPINAS
GVDAAKALSLLKSSRLENLIAQLPGKEKKNGLFLGLNALSGLTNPFSNFDLPAILDKLQSLKDTRYDDDLNLLAQIGDQYADLFAAKNLSDLILLDLIRNEITKAPLASISMRKYDEHQLITLLILKVR
QQLFPEKYEIFDFFQSKNMYAGGQTLKSYEPPFYIKFLKEMDGTVEELVLKINREDLRLRRKQTFDNG
SPIPHQIHLHAILRQRQEDYYFPLKDRNREIKELTRIFIPYYVPLARNSRFWAMTRKSEETITPF
NFEENWDGKASAQSFIERMTNFDKNPSQELKVPSLKKLLSYETFVYNETLKTVYKGMPKLSGFSEM
KAILDLFLKTNRKVTKVQKEDFYKIEFDSVEISGVEDRFNASLGTYHDLKIIKDKFDLNEEDE
DILEDIVLFLEDREMYAHLOFTDKVMQLLPFRRYTGWGRSRLKINGIRDSQKSTIIL
DFLKSDGQANRMRFSQILHDDSITFKIDQAKOSQGQLSLHEHIANLPSAUIGKIOLTVQKVDDEL
KVGRRHKPENIVIEMARENQYTTQGKQNERSMRKIEEIKELGSQILKEHPVENTQLONEKYLLYL
QNGRMDYMVDQELDNLSSDLVDHPIDNKLDSIDNKLCDTRSNKGSVDPEEVEKMKMNH
QLNNAKLLITQRFDNLTKAERGKSDLKAGFIKRQOLVETRQKTHAVQILDLSMTNKFYDENDLIERK
VKVTILSKLVSDFKDQFYKVEEINNYHAHDAYLAENVNG CLEAGTRIDFPTVGTREIDIHugh Amy
GPQGSGSLALSITADQVMSALLDAEFFPYSEYDTSFSPEASMMMGLLTLNADRELVMINAWKRPV
GFVDLTHQDLHERALWELIMIGLWRSMEPHPKLFAJPNLLDNRONGQGVEGMRVEFDMLLASS
RFRMNNQLEGEVFCISKI ILLNSGVITFLSLKLSEKEDHIHALKDTILDILHARGLTQLQHQC
QHRLAQLILLSHIRSNKSGMEHSLSKYWNPLYLIMELDAHRLHAGGASGARQAPADALDDK
FLHMLAEGLNYSVIREPLTRARDFLEVELHTLAEEVWNHCALIJKLMNOPSEFYSVGYKY
DVRVMIAKSEQIEKFKYSNIMNFETEITLANEEIRKPLEIETNGTEIWEVDKGRDFATVR
KVLSMPVQNVITEQVGFGSKSERIKRNRSKIARDDKWKPGDFSPTAVSVLWAKVEKG
KSSKLKSVKELLGITIMERSPEKNPQIDLEAKGKVEKKDLI IKLPSYLSELENGRKRMLASGAGEL
QKGKNAELPSKYVFNLYLASHEYKLSGSPDNEQKQLFVEQHKYXDEIEIQISEFPSKVILADANL
KVLASYNKRHKDPIREQAENI IHLFTLTNGAPAAFYDFTIDRKYRSTKEVLDATL IHQS ITGLY
ETRIDLSQGDL

SEQ ID NO: 36

SEQ ID NO: 37
DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFFDQSKNGAYIDGGASQEEFYFKIFIPILEMDGTEELLVKLNREDLLRQRTFDNG
SIPHQIHGGHLLRAKQDEYFPPFDKMKKQLRRKYRRTGWRSLRKLINGRDSQKSTIT
DPLSFDDANIFNLQGQIHDSDLFTKEDIQAAVSGQGDSLEHIANLAPSAIKGILQTVKVDVDEL
KVGMRRHPENIEMARENTQTPQGKQNSRMRKIEEIGKLSIQHEPLEVTNQLKELTYYLYL
QRGRMYQDQDLNRLSDKFTDHPQFSPFQDSDKSDINVLTSDKRRGSDNVPSEEEVKKMKNYR
QNLAKLITQKFRDNLTAERGSSLDELDKAFQKRIQLVEQITKVQIAOLDSNMTYNDKENDLRE
VKVITLKSVLSDIFKDFQFYKVREINNYHIHAADYLANAVGTAFLIKYPKLESEFYGDYKYVYDRK
MIASEQEIGKATAYFYFSINMNFKTEITILANGAEIRKRFLEITNGETGEIWVDGRDFPAVRKVLK
MPQVNIKVTEVQTPGGFSESILPKRNSKLIARKKDWDKFGGGEDSFPTAVAYSVLVAVKVEKGSKK
LKLCAEGTIFDVGTHIREDWDFKGLQKIQVAAKGDTLALFPRSWFDQGTRDVDGIARAIAGAII
VWATPDHKVLITYEGWARGAAGELLRGDVQAGPSGSGLSALSLTDQAMSALLADEPFIYLSEYDTPSF
SEASMMGLLTNLAERDLRELHMINWAKARPVFVDDLHDLQAHLLERAWELIMIGLWRSMHEFKLLFA
PNLILDQRAVQIFMVPDMLLSTHRMNLQIEEVEFCVLCKIIINSLLSHYFVTLSSKLEEKEDHRLRADK
ITDLILHLMAKAGLSTQQQHRQRAQLTLILISHRHRSMSNGMELYSMYKWNPLYIILL
EMDLARHLAGGGSARVQFADALDDFKLMAELGRLSYVIRELPTRARTDFLEVELHTLVAE
GVWNC_KVELLITSHMSEEKFSPDFLEAKGYEVKKDDILKLPKYLFELENGRKRMLASAGELQK
GNEALSFLSKYVNFLYSLKGSQDNQKLFVQHEQKLHLYDEIIEQISEFSKRVILANDLO
KVLSAYNHKRDPIREQAENIHLFPTLNLGAPAFAKDFDTIDRKYRSTKETVLDATLTHQSTITY
ETRIDLSSLQGD (SEQ ID NO:40)

Cas9:Intein (37R3-2; in double underline) replacing Serl274

MDKYSIGLDIGTNSGAVWAVITQDYEKVPSKFKVLGNTDSIHKKNLGALFLDEGETAGAELRTKTT
RRYTRKRRNCRYLQIEFQIFHRRLEELFSLVEEDKKKHERHPIGN1VDEAYE8KPYT5Y
HLRKLVDSTDKADLRILYLAHALMKFRGHFLIEGODLNPNSVDKLFILQVQNYQFLEENPNS
GVDAKAILSRLKSSRERENIALQPLGEKKNGLFGMLASLGLTPNFKSNFPLADEALKLSKDTYD
DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVTQKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRQRTFDNG
SIPHQIHGGHLLRAKQDEYFPPFDKMKKQLRRKYRRTGWRSLRKLINGRDSQKSTIT
DPLSFDDANIFNLQGQIHDSDLFTKEDIQAAVSGQGDSLEHIANLAPSAIKGILQTVKVDVDEL
KVGMRRHPENIEMARENTQTPQGKQNSRMRKIEEIGKLSIQHEPLEVTNQLKELTYYLYL
QRGRMYQDQDLNRLSDKFTDHPQFSPFQDSDKSDINVLTSDKRRGSDNVPSEEEVKKMKNYR
QNLAKLITQKFRDNLTAERGSSLDELDKAFQKRIQLVEQITKVQIAOLDSNMTYNDKENDLRE
VKVITLKSVLSDIFKDFQFYKVREINNYHIHAADYLANAVGTAFLIKYPKLESEFYGDYKYVYDRK
MIASEQEIGKATAYFYFSINMNFKTEITILANGAEIRKRFLEITNGETGEIWVDGRDFPAVRKVLK
MPQVNIKVTEVQTPGGFSESILPKRNSKLIARKKDWDKFGGGEDSFPTAVAYSVLVAVKVEKGSKK
LKLCAEGTIFDVGTHIREDWDFKGLQKIQVAAKGDTLALFPRSWFDQGTRDVDGIARAIAGAII
VWATPDHKVLITYEGWARGAAGELLRGDVQAGPSGSGLSALSLTDQAMSALLADEPFIYLSEYDTPSF
SEASMMGLLTNLAERDLRELHMINWAKARPVFVDDLHDLQAHLLERAWELIMIGLWRSMHEFKLLFA
PNLILDQRAVQIFMVPDMLLSTHRMNLQIEEVEFCVLCKIIINSLLSHYFVTLSSKLEEKEDHRLRADK
ITDLILHLMAKAGLSTQQQHRQRAQLTLILISHRHRSMSNGMELYSMYKWNPLYIILL
EMDLARHLAGGGSARVQFADALDDFKLMAELGRLSYVIRELPTRARTDFLEVELHTLVAE
GVWNC_KVELLITSHMSEEKFSPDFLEAKGYEVKKDDILKLPKYLFELENGRKRMLASAGELQK
GNEALSFLSKYVNFLYSLKGSQDNQKLFVQHEQKLHLYDEIIEQISEFSKRVILANDLO
KVLSAYNHKRDPIREQAENIHLFPTLNLGAPAFAKDFDTIDRKYRSTKETVLDATLTHQSTITY
ETRIDLSSLQGD (SEQ ID NO:40)
In some embodiments, the intein inserted into the Cas9 protein is ligand-dependent. In some embodiments, the ligand-dependent inteins comprise a modified ligand-binding domain of the estrogen receptor protein, embedded into a modified RecA intein from *M. tuberculosis*. In some embodiments, the ligand-binding domain is derived from an estrogen receptor protein, for example, from the human estrogen receptor. The sequence of the human estrogen receptor and the location of the ligand-binding domain within the human estrogen receptor are known to those of skill in the art. Non-limiting, exemplary sequences of the human estrogen receptor can be retrieved from RefSeq database entries NP_0001 16 (isoform 1); NP_001 116212 (isoform 2); NP_001 116213 (isoform 3); and NP_001 116214 (isoform 4) from the National Center for Biotechnology Information (NCBI, www.ncbi.nlm.nih.gov). In some embodiments, the ligand-binding domain of a ligand-dependent intein provided herein comprises or is derived from a sequence comprising amino acid residues 304-551 of the human estrogen receptor.

It will be appreciated by those of skill in the art that other ligand-dependent inteins are also suitable and useful in connection with the Cas9 proteins and methods provided herein. For example, some aspects of this invention provide Cas9 proteins comprising ligand-dependent inteins that comprise a ligand-binding domain of a hormone-binding protein, *e.g.*, of an androgen receptor, an estrogen receptor, an ecdysone receptor, a glucocorticoid receptor, a mineralocorticoid receptor, a progesterone receptor, a retinoic acid receptor, or a thyroid hormone receptor protein. Ligand-binding domains of hormone-binding receptors, inducible fusion proteins comprising such ligand-binding domains, and methods for the generation of such fusion proteins are known to those of skill in the art (see, *e.g.*, Becker, D., Hollenberg, S., and Ricciardi, R. (1989). Fusion of adenovirus E1A to the glucocorticoid receptor by high-resolution deletion cloning creates a hormonally inducible viral transactivator. Mol. Cell. Biol. 9, 3878-3887; Boehmelt, G., Walker, A., Kabrun, N., Mellitzer, G., Beug, H., Zanke, M., and Enrietto, P. J. (1992). Hormone-regulated v-rel estrogen receptor fusion protein: reversible induction of cell transformation and cellular gene expression. EMBO J 11, 4641-4652; Braselmann, S., Graninger, P., and Busslinger, M. (1993). A selective transcriptional induction system for mammalian cells based on Gal4-estrogen receptor fusion proteins. Proc Natl Acad Sci U S A 90, 1657-1661; Furga G, Busslinger M (1992). Identification of Fos target genes by the use of selective induction systems. J. Cell Sci. Suppl 16,97-109; Christopherson, K. S., Mark, M. R., Bajaj, V., and Godowski, P. J. (1992). Ecdysteroid-dependent regulation of genes in mammalian cells by a Drosophila ecdysone receptor and chimeric transactivators. Proc Natl Acad Sci U S A 89,

[0083] Additional exemplary inteins, ligand-binding domains, and ligands suitable for use in the Cas9 proteins provided herein are described in International Patent Application, PCT/US20 12/028435, entitled "Small Molecule-Dependent Inteins and Uses Thereof," filed March 9, 2012, and published as WO 2012/125445 on September 20, 2012, the entire contents of which are incorporated herein by reference. Additional suitable inteins, ligand-binding domains, and ligands will be apparent to the skilled artisan based on this disclosure.

[0084] The ligand-dependent inteins provided herein are inactive (or only minimally active) in the absence of the appropriate ligand, but can be induced to be active, and, thus, to self-excise, by contacting them with a ligand that binds the ligand-binding domain of the human estrogen receptor. Small molecule ligands binding the ligand-binding domain of the estrogen receptor (e.g., the human estrogen receptor), and thus useful to induce the activity of the ligand-dependent inteins described herein, are known to those of skill in the art. In some embodiments, the ligand used to induce the activity of the ligand-dependent inteins described herein specifically binds to the ligand-binding domain of the estrogen receptor. In some embodiments, the ligand binds the ligand-binding domain of a ligand-dependent intein.
provided herein with high affinity, for example, with an affinity of at least about $10^{-10}$ M, at least about $10^{-9}$ M, at least about 10 M, at least about 10 M, at least about 10 M, or at least about 10 M. Examples of appropriate estrogen receptor-binding ligands that are useful to induce the activity of the ligand-dependent inteins provided herein, for example, the ligand-dependent inteins provided in SEQ ID NOs 3-8, include, but are not limited to, 17β-estradiol, 17α-ethynyl estradiol, tamoxifen and tamoxifen analogs (e.g., 4-hydroxytamoxifen (4-HT, 4-OHT), 3-hydroxytamoxifen (drolotixifene)), tamoxifen metabolites (e.g., hydroxytamoxifen, endoxifen), raloxifene, toremifene, ICI-182, and ICI-780. Other useful ligands will be apparent to those of skill in the art, and the invention is not limited in this respect.

In some embodiments, any of the Cas9 proteins comprising inteins (e.g., SEQ ID NOs:27-41) can be modified so as to generate a Cas9 nickase comprising an intein (e.g., by making one of a D10A or H840A mutation relative to the Cas9 sequence lacking an intein), or to generate a dCas9 protein comprising an intein (e.g., by making both D10A and H840A mutations relative to the Cas9 sequence lacking an intein). In some embodiments, any of the Cas9 proteins comprising inteins (e.g., SEQ ID NOs:27-41) have additional features, for example, one or more linker sequences, localization sequences, such as nuclear localization sequences (NLS; e.g., MAPKKKRKVGRIHRGVP (SEQ ID NO:42)); cytoplasmic localization sequences; export sequences, such as nuclear export sequences; or other localization sequences, as well as sequence tags that are useful for solubilization, purification, or detection of the fusion proteins. Suitable localization signal sequences and sequences of protein tags are provided herein and are known in the art, and include, but are not limited to, biotin carboxylase carrier protein (BCCP) tags, myc-tags, calmodulin-tags, FLAG-tags (e.g., 3xFLAG TAG: MDYKHDGDGYKDDDDK (SEQ ID NO:43)), hemagglutinin (HA) tags, polyhistidine tags, also referred to as histidine tags or His-tags, maltose binding protein (MBP)-tags, nus-tags, glutathione-S-transferase (GST) tags, green fluorescent protein (GFP) tags, thioredoxin-tags, S-tags, Softags (e.g., Softag 1, Softag 3), strep-tags, biotin ligase tags, FIAAS tags, V5 tags, and SBP-tags. Additional suitable sequences will be apparent to those of skill in the art.

In some embodiments, ligand-dependent site-specific enzymes (e.g., fusion proteins) are provided which comprise a Cas9 variant (e.g., dCas9), a ligand-dependent intein, and one or more other polypeptide domains having a particular enzymatic activity. In some embodiments, the fusion protein comprises a nuclease inactivated Cas9 domain (e.g., dCas9), wherein the dCas9 domain comprises an intein sequence inserted in place of or in
addition to any amino acid in dCas9. In some embodiments the inserted intein sequence replaces Cys80, Ala27, Thr46, Ser219, Thr333, Thr519, Cys574, Thr622, Ser701, Ala728, Thr995, Serl006, Serll54, Serll59, or Serl274 of dCas9 (SEQ ID NO:5). In some embodiments, the inserted intein sequence replaces Alal27, Thr46, Ser219, Thr519, or Cys574 of dCas9 (SEQ ID NO:5). In some embodiments, the intein is inserted into another domain of the fusion protein (i.e., not in the Cas9 domain, e.g., not in the dCas9 domain), such as the domain having a particular enzymatic activity. In some embodiments, the domain having a particular enzymatic activity is a nuclease domain (e.g., FokI), a recombinase catalytic domain (e.g., Hin, Gin, or Tn3 recombinase domains), a nucleic acid-editing domain (e.g., a deaminase domain), a transcriptional activator domain (e.g., VP64, p65), a transcriptional repressor domain (e.g., KRAB, SID), or an epigenetic modifier (e.g., LSD1 histone demethylase, TET1 hydroxylase). The intein that is inserted into the fusion protein can be any ligand-dependent intein, e.g., those described herein. For example, in some embodiments, the intein that is inserted into a Cas9 protein comprises in part or in whole, a sequence that is at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% percent identical to any one of SEQ ID NO:7-14.

In some embodiments, the general architecture of exemplary fusion proteins provided herein comprises the structure:

[NH$_2$]-[enzymatic domain]-[dCas9]-[COOH] or
[NH$_3$]-[dCas9]-[enzymatic domain]-[COOH]:

wherein NH$_2$ is the N-terminus of the fusion protein, COOH is the C-terminus of the fusion protein, dCas9 comprises an intein as provided herein, and the enzymatic domain comprises a nuclease domain (e.g., FokI), a recombinase catalytic domain (e.g., Hin, Gin, or Tn3 recombinase domains), a nucleic acid-editing domain (e.g., a deaminase domain), a transcriptional activator domain (e.g., VP64, p65), a transcriptional repressor domain (e.g., KRAB, SID), or an epigenetic modifier (e.g., LSD1 histone demethylase, TET1 hydroxylase).

In some embodiments, the intein is comprised in a domain other than dCas9 (e.g., in an enzymatic domain), or is located between two domains.

Additional features may be present, for example, one or more linker sequences between certain domains. Other exemplary features that may be present are localization sequences, such as nuclear localization sequences (NLS; e.g., MAPKKKRKVGIHRGVP (SEQ ID NO:42)); cytoplasmic localization sequences; export sequences, such as nuclear export sequences; or other localization sequences, as well as sequence tags that are useful for solubilization, purification, or detection of the fusion proteins. Suitable localization signal
sequences and sequences of protein tags are provided herein and are known in the art, and include, but are not limited to, biotin carboxylase carrier protein (BCCP) tags, myc-tags, calmodulin-tags, FLAG-tags (e.g., 3xFLAG TAG: MDYKDHGDYKDHDYKDDDDK (SEQ ID NO:43)), hemagglutinin (HA) tags, polyhistidine tags, also referred to as histidine tags or His-tags, maltose binding protein (MBP)-tags, nus-tags, glutathione-S-transferase (GST) tags, green fluorescent protein (GFP) tags, thioredoxin-tags, S-tags, Softags (e.g., Softag 1, Softag 3), strep-tags, biotin ligase tags, FlAsH tags, V5 tags, and SBP-tags. Additional suitable sequences will be apparent to those of skill in the art.

[0089] In some embodiments, the enzymatic domain comprises a nuclease or a catalytic domain thereof. For example, in some embodiments, the general architecture of exemplary ligand-dependent dCas9 fusion proteins with a nuclease domain comprises the structure:

\[
[NH_2]-[NLS]-[dCas9]-[nuclease]-[COOH],
\]

\[
[NH_2]-[NLS]-[nuclease]-[dCas9]-[COOH],
\]

\[
[NH_2]-[dCas9]-[nuclease]-[COOH],
\]

\[
[NH_2]-[nuclease]-[dCas9]-[COOH];
\]

wherein NLS is a nuclear localization signal, dCas9 comprises an intein as provided herein, NH\textsubscript{2} is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein. In some embodiments, a linker is inserted between the dCas9 and the nuclease domain. In some embodiments, a linker is inserted between the NLS and the nuclease and/or dCas9 domain. In some embodiments, the NLS is located C-terminal of the nuclease and/or the dCas9 domain. In some embodiments, the NLS is located between the nuclease and the dCas9 domain. Additional features, such as sequence tags, may also be present. In some aspects, the nuclease domain is a nuclease requiring dimerization (e.g., the coming together of two monomers of the nuclease) in order to cleave a target nucleic acid (e.g., DNA). In some embodiments, the nuclease domain is a monomer of the Fokl DNA cleavage domain. The Fokl DNA cleavage domain is known, and in some aspects corresponds to amino acids 388-583 of Fokl (NCBI accession number J04623). In some embodiments, the Fokl DNA cleavage domain corresponds to amino acids 300-583, 320-583, 340-583, or 360-583 of Fokl. See also Wah et al, "Structure of Fokl has implications for DNA cleavage" *Proc. Natl. Acad. Sci. USA.* 1998; 1;95(18): 10564-9; Li et al, "TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and Fokl DNA-cleavage domain" *Nucleic Acids Res.* 2011; 39(1):359-72; Kim et al, "Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain" *Proc. Natl Acad. Sci. USA.* 1996; 93:1156-1160; the entire contents of each are
herein incorporated by reference). In some embodiments, the FokI DNA cleavage domain corresponds to, or comprises in part or whole, the amino acid sequence set forth as SEQ ID NO:6. In some embodiments, the FokI DNA cleavage domain is a variant of FokI (e.g., a variant of SEQ ID NO:6), as described herein. Other exemplary compositions and methods of using dCas9-nuclease fusion proteins can be found in U.S. patent application U.S.S.N 14/320,498; titled "Cas9-FokI fusion Proteins and Uses Thereof," filed June 30, 2014; the entire contents of which are incorporated herein by reference.

[0090] In some embodiments, the enzymatic domain comprises a recombinase or other catalytic domain thereof. For example, in some embodiments, the general architecture of exemplary ligand-dependent dCas9 fusion proteins with a recombinase domain comprises the structure:

\[
[NH_2]-[NLS]-[dCas9]-[recombinase]-[COOH],
\]

\[
[NH_2]-[NLS]-[dCas9]-[recombinase]-[COOH],
\]

\[
[NH_2]-[dCas9]-[recombinase]-[COOH],
\]

\[
[NH_2]-[recombinase]-[dCas9]-[COOH];
\]

wherein NLS is a nuclear localization signal, dCas9 comprises an intein as provided herein, NH₂ is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein. In some embodiments, a linker is inserted between the dCas9 and the recombinase domain. In some embodiments, a linker is inserted between the NLS and the recombinase and/or dCas9 domain. In some embodiments, the NLS is located C-terminal of the recombinase domain and/or the dCas9 domain. In some embodiments, the NLS is located between the recombinase domain and the dCas9 domain. Additional features, such as sequence tags, may also be present. By "catalytic domain of a recombinase," it is meant that a fusion protein includes a domain comprising an amino acid sequence of (e.g., derived from) a recombinase, such that the domain is sufficient to induce recombination when contacted with a target nucleic acid (either alone or with additional factors including other recombinase catalytic domains which may or may not form part of the fusion protein). In some embodiments, a catalytic domain of a recombinase does not include the DNA binding domain of the recombinase. In some embodiments, the catalytic domain of a recombinase includes part or all of a recombinase, e.g., the catalytic domain may include a recombinase domain and a DNA binding domain, or parts thereof, or the catalytic domain may include a recombinase domain and a DNA binding domain that is mutated or truncated to abolish DNA binding activity. Recombinases and catalytic domains of recombinases are known to those of skill in the art, and include, for example, those described herein. In some embodiments, the
catalytic domain is derived from any recombinase. In some embodiments, the recombinase catalytic domain is a catalytic domain of aTn3 resolvase, a Hin recombinase, or a Gin recombinase. In some embodiments, the catalytic domain comprises a Tn3 resolvase (e.g., Stark Tn3 recombinase) that is encoded by a nucleotide sequence comprising, in part or in whole, SEQ ID NO:44, as provided below. In some embodiments, a Tn3 catalytic domain is encoded by a variant of SEQ ID NO:44. In some embodiments, a Tn3 catalytic domain is encoded by a polynucleotide (or a variant thereof) that encodes the polypeptide corresponding to SEQ ID NO:45. In some embodiments, the catalytic domain comprises a Hin recombinase that is encoded by a nucleotide sequence comprising, in part or in whole, SEQ ID NO:46, as provided below. In some embodiments, a Hin catalytic domain is encoded by a variant of SEQ ID NO:46. In some embodiments, a Hin catalytic domain is encoded by a polynucleotide (or a variant thereof) that encodes the polypeptide corresponding to SEQ ID NO:47. In some embodiments, the catalytic domain comprises a Gin recombinase (e.g., Gin beta recombinase) that is encoded by a nucleotide sequence comprising, in part or in whole, SEQ ID NO:48, as provided below. In some embodiments, a Gin catalytic domain is encoded by a variant of SEQ ID NO:48. In some embodiments, a Gin catalytic domain is encoded by a polynucleotide (or a variant thereof) that encodes the polypeptide corresponding to SEQ ID NO:49. Other exemplary compositions and methods of using dCas9-recombinase fusion proteins can be found in U.S. patent application U.S.S.N 14/320,467; titled "Cas9 Variants and Uses Thereof," filed June 30, 2014; the entire contents of which are incorporated herein by reference.

Stark Tn3 recombinase (nucleotide: SEQ ID NO:44; amino acid: SEQ ID NO:45):

```
ATGGGCCCCGCTTTGCGTCCAGCAAGGGCTGGTTCACACACAGTCATCATGTTTGCAAAGTGGACGGGCTTAAAGAT
GCCAGATGAGGGACGGCAAGCAGCTAAGTTGAAAGGTATCAAATTTGGCAGACGAAGG
```

Hin Recombinase (nucleotide: SEQ ID NO:46; amino acid: SEQ ID NO:47):

```
ATGGGACCATTAAGGTGTCCATACATCGACCGCAGACCACGCTATCTGATCTCCACG
GCCGACCTGGCAGATCTCATCTCCATACATCGACCGCAGACCACGCTATCTGATCTCCACG
```

15/320,467; titled "Cas9 Variants and Uses Thereof," filed June 30, 2014; the entire contents of which are incorporated herein by reference.

Stark Tn3 recombinase (nucleotide: SEQ ID NO:44; amino acid: SEQ ID NO:45):

```
ATGGGCCCCGCTTTGCGTCCAGCAAGGGCTGGTTCACACACAGTCATCATGTTTGCAAAGTGGACGGGCTTAAAGAT
GCCAGATGAGGGACGGCAAGCAGCTAAGTTGAAAGGTATCAAATTTGGCAGACGAAGG
```

Hin Recombinase (nucleotide: SEQ ID NO:46; amino acid: SEQ ID NO:47):

```
ATGGGACCATTAAGGTGTCCATACATCGACCGCAGACCACGCTATCTGATCTCCACG
GCCGACCTGGCAGATCTCATCTCCATACATCGACCGCAGACCACGCTATCTGATCTCCACG
```

15/320,467; titled "Cas9 Variants and Uses Thereof," filed June 30, 2014; the entire contents of which are incorporated herein by reference.

Stark Tn3 recombinase (nucleotide: SEQ ID NO:44; amino acid: SEQ ID NO:45):

```
ATGGGCCCCGCTTTGCGTCCAGCAAGGGCTGGTTCACACACAGTCATCATGTTTGCAAAGTGGACGGGCTTAAAGAT
GCCAGATGAGGGACGGCAAGCAGCTAAGTTGAAAGGTATCAAATTTGGCAGACGAAGG
```

Hin Recombinase (nucleotide: SEQ ID NO:46; amino acid: SEQ ID NO:47):

```
ATGGGACCATTAAGGTGTCCATACATCGACCGCAGACCACGCTATCTGATCTCCACG
GCCGACCTGGCAGATCTCATCTCCATACATCGACCGCAGACCACGCTATCTGATCTCCACG
```

15/320,467; titled "Cas9 Variants and Uses Thereof," filed June 30, 2014; the entire contents of which are incorporated herein by reference.
ATGGGCCGATTCCTTTTACGTCATGTCCCCTCGCTGAAATGGAGCGCGAACTTATTGTTGAACGGACTTTGGCTGGACTGGCAGCGGCTAGAGCACAGGGCCGACTTGGA (SEQ ID NO:46)

MATIGYRST V 1 DQ811 DLQRNLTSANDDRFEDRI SGKIANPGLKRALKVYNKDITLVLWVLGLRSLV AL1 SELHERGAHFHSTDS 1 DTS SAMGRFFFYVMSALAEMEREL IVERTLAAARAQGRLG (SEQ ID NO:47)

Gin beta recombinase (nucleotide: SEQ ID NO:48; amino acid: SEQ ID NO:49):

ATGCTCATTGGCTATGTAAGGGTCAGCACCAATGACCAAAACACAGACTTGCAACGCAATGCTTTGGTGCCGGATGTGAACAGATATTTGAAGATAAACTGAGCGGCACTCGGACAGACAGACCTGGGCTTAAGAGAGCACTGAAAAGACTGCAGAAGGGGGACACCCTGGTCGTCTGGAAACTGGATCGCCTCGGACGCAGCATGAAACATCTGATTAGCCTGGTTGGTGAGCTTAGGGAGAGAGGAATCAACTTCAGAAGCCTGACCGACTCCATCGACACCAGTAGCCCCATGGGACGATTCTTCTTTTATGTGATGGGAGCACTTGCTGAGATGGAAAGAGAGCTTATTATCGAAAGAACTATGGCTGGTATCGCTGCTGCCCGGAACAAAGGCAGACGGTTCGGCAGACCGCCGAAGAGCGGC (SEQ ID NO:48)

ML IGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSTGDTRPGLKRALKRLQKGDITLVLWVLGLRSMKHL 1 S LVGELRERGINFRSTDS 1 DTS SPGRFFFYVMGALAEEREL IERTMAGIAARNGRFRGPPKSG (SEQ ID NO:49)

[0091] In some embodiments, the enzymatic domain comprises a deaminase or a catalytic domain thereof. For example, in some embodiments, the general architecture of exemplary dCas9 fusion proteins with a deaminase enzyme or domain comprises the structure:

[NH2]-[NLS]-[Cas9]-[deaminase]-[COOH],
[NH2]-[NLS]-[deaminase]-[Cas9]-[COOH],
[NH2]-[Cas9]-[deaminase]-[COOH], or
[NIL]-[deaminase]-[Cas9]-[COOH];

wherein NLS is a nuclear localization signal, dCas9 comprises an intein as provided herein, NH2 is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein. In some embodiments, a linker is inserted between the dCas9 and the deaminase domain. In some embodiments, a linker is inserted between the NLS and the deaminase and/or dCas9 domain. In some embodiments, the NLS is located C-terminal of the deaminase and/or the dCas9 domain. In some embodiments, the NLS is located between the deaminase domain and the dCas9 domain. Additional features, such as sequence tags, may also be present. One exemplary suitable type of nucleic acid-editing enzymes and domains are cytosine deaminases, for example, of the apolipoprotein B miRNA-editing complex (APOBEC) family of cytosine deaminase enzymes, including activation-induced cytidine deaminase (AID) and apolipoprotein B editing complex 3 (APOBEC3) enzyme. Another exemplary suitable type of nucleic acid-editing enzyme and domain thereof suitable for use in the present invention include adenosine deaminases. For example, an ADAT family
adenosine deaminase can be fused to a dCas9 domain comprising an intein. Some exemplary suitable nucleic-acid editing enzymes and domains, e.g., deaminases and deaminase domains, that can be fused to dCas9 domains comprising inteins according to aspects of this disclosure are provided below. It will be understood that, in some embodiments, the active domain of the respective sequence can be used, e.g., the domain without a localizing signal (nuclear localizing signal, without nuclear export signal, cytoplasmic localizing signal). Other exemplary compositions and methods of using dCas9-nuclease fusion proteins can be found in U.S. patent application U.S.S.N. 14/325,815; titled "Fusions of Cas9 Domains and Nucleic Acid-Editing Domains," filed July 8, 2014; the entire contents of which are incorporated herein by reference.

**Human AID:**

```
<table>
<thead>
<tr>
<th>Sequence</th>
<th>nuclear localization signal</th>
<th>nuclear export signal</th>
</tr>
</thead>
</table>
| MDSLLMRKFLYQFKNVRAKGRHETYL | YVVKRDRDTSFSLDFGYLRNKNGCHVELLFLRY | SDWDLDPGR
| YRVTWFTSWSPCYDCAHVFADFLRNPNSLRI | TFARTLYFCEKRKAEPGLRRLHRAGVQIAIMTFKDFYFCWNT | FVENHRTFKAWEGLHENVSRQLRRILLLP | LYEVDLRDARFLGL (SEQ ID NO:50) |
```

**Mouse AID:**

```
<table>
<thead>
<tr>
<th>Sequence</th>
<th>nuclear localization signal</th>
<th>nuclear export signal</th>
</tr>
</thead>
</table>
| MPSLLMKQKFLYHFKNNVARWAKGHR | YVVKRDRDTSFSLDFGYLRNKNGCHVELLFLRY | SDWDLDPGR
| YRVTWFTSWSPCYDCAHVFADFLRNPNSLRI | TFARTLYFCEKRKAEPGLRRLHRAGVQIAIMTFKDFYFCWNT | FVENREKTFKAWEGLHENVSRQLRRILLLP | LYEVPPLRPAFRTLGL (SEQ ID NO:51) |
```

**Dog AID:**

```
<table>
<thead>
<tr>
<th>Sequence</th>
<th>nuclear localization signal</th>
<th>nuclear export signal</th>
</tr>
</thead>
</table>
| MPSLLMKQRKFLYHFKNNVARWAKGHR | YVVKRDRDTSFSLDFGYLRNKNGCHVELLFLRY | SDWDLDPGR
| YRVTWFTSWSPCYDCAHVFADFLRNPNSLRI | TFARTLYFCEKRKAEPGLRRLHRAGVQIAIMTFKDFYFCWNT | FVENREKTFKAWEGLHENVSRQLRRILLLP | LYEVPPLRPAFRTLGL (SEQ ID NO:52) |
```

**Bovine AID:**

```
<table>
<thead>
<tr>
<th>Sequence</th>
<th>nuclear localization signal</th>
<th>nuclear export signal</th>
</tr>
</thead>
</table>
| MPSLLKQKORQFLYQFKNVRAKGRHETYL | YVVKRDRDTSFSLDFGYLRNKNGCHVELLFLRY | SDWDLDPGR
| YRVTWFTSWSPCYDCAHVFADFLRNPNSLRI | TFARTLYFCEKRKAEPGLRRLHRAGVQIAIMTFKDFYFCWNT | FVENREKTFKAWEGLHENVSRQLRRILLLP | LYEVPPLRPAFRTLGL (SEQ ID NO:53) |
```

**Mouse APOBEC-3:**

```
<table>
<thead>
<tr>
<th>Sequence</th>
<th>nuclear localization signal</th>
<th>nuclear export signal</th>
</tr>
</thead>
</table>
| MGFPLGCSHRKCYSPIRNLI | SQETFKFHKNLYAKRKPTFLCYEVTRKCPSVFSLHGVFKNPN | HAEI
| CYFLYWHPKVKLSPREEFKITYWMSWSPCFEC | AEQIVRLATHHNLSPIFFSSLRYNQVQETQNLCRLVQEG | HAEILFLPKIRSM
| AQVAMPYFEKCKWKKVGPNGGRFRPKRILLTNRYPQSKLQELRCPYVPSSSHSTLSNICLTKLPETR | FCVEGRMPPLSEEEFQYFQNYQRVHKLCYHHRMKPFLCYQLEQFNQAPLKGCLLSEKGQ |
Rat APOBEC-3:

MGPFCLGCSHRKCYSPIRNL SQETFKHFKNLRYAI DRKDTFLCYEVTRKDCSPVSLHGHGFKNKDNH HAE1C FLYWFHDKVLKVLSPREEFKI TWYMSSWPCFECAEQVLRFLAHTHNLSDLIFS SRLYNIRDPEQNOQLCRLVQEG A0VAAMDLYEFKKCKWKFVDNGGRRFRPWKLLTNFRYQDSKLEI LRPCYIPVPS S S STLNSNCITKGLPETR FCVERRRVHLSEEFYFSQFPQNYQVHKLCCYHYGKYPYLCYQEQFNGQAPLKGCLLSEKQG HAE1FLLDKIRSM ELSQVI I TCYLTWSPCPNC AWQALAFKRDPRDL I LHIYTSRLYFHWKRPFQKGLCSLWQSGILVDVMDLPQFTDC WTNFVNPKFPRF PWKGLIE I SRRTQRRLRRIKESWGLQDLNDFGNLQGPPMS (SEQ ID NO:55)
(underline: nucleic acid editing domain)

Rhesus macaque APOBEC-3G:

MVEPMDPRTFVSNFNRP LLSGLNTVWLCCEVTKDPSGPPLDAKIFQGKYSKAYHPG "RFIRWFHKWROLAH DQEYKVT" TVSVSPCiriCANSVATFTLAOAKPKVLT IFVARLYYFVKPDYQQALRI LCQKRGGPHATMKIMYNENF QDCKWKFVDRSGKPFKRPNRLPKHYTTLQATLGELLRHMDPGTSTNFPKVPWSQGQHETLYCFLVERLHNDTW VPLNQRFLRLQAPNHI GFPGKRHAELCDFLIPFIVCLDGeeriRyrCnrSIVSPCFCSAQEMAKF 1 SNEHVSILC IFAARI YDDQGRGYEQEGRLR HDGAKIAMMNYESFECWDTFVDRQGRFQPWPDLDEHSQALSGLRAI (SEQ ID NO:56)
(bold italic: nucleic acid editing domain; underline: cytoplasmic localization signal)

Chimpanzee APOBEC-3G:

MKPHFRNPVERMYODTSDFNYFNRP LLSHRNTVWLCEYVTKGPRSPRPDLADIFRGOVYSKLKYHPEMRFHWF __ KWRKLRHDKQEVETVEYW SWSPTCKETRDVATFLADPKVLT IFVARLYYFWDYQFQALRSLCQKRDGPRATMK IMNYDEFCWKSFGYQSFRLFEPWNRLPKYII LLLHIMLGE 1 LHRSDPPFSTNFPKVPWSQGQHETLYCFLVERLHNDTW LNLQRRGFLCNQAPHI KgfLEGK HAE1CFLDVFIPWFLDLHQDQYRVCFTSTWSPCFC __AQEMAKF1S NNKHVSILC IFAARI YDDQGRGQEGRLR HDGAKIAMMNYESFECWDTFVDRQGRFQPWPDLDEHSQALSGLRAI (SEQ ID NO:57)
(underline: nucleic acid editing domain; double underline: cytoplasmic localization signal)

Green monkey APOBEC-3G:

MNPQINMQEPMDDFYVFYFNRP LLSGRNTVWLCCEVTKDPSGPPDLANIFQGKLYPEAKDHPEMFKFLHWFR__ KWRKLRHDKQEVETVEYW SWSPTCRTCANSVATFTLAEDPKVLT IFVARLYYFWDYQQALR1 LCQERGGPHATMK IMNYEFCWHCWFVDFQGKFKPFKNLPHYTTLHATLGERHHMDPGTSTNFPKVPWSQGQHETLYCYKVE RSHNHTWVILLNQRHFLNQAPDRHGFPGKRHAELCDFL __IFWFLKLDQQYRVTCFTSTWSPCFC __AQEMAKF1S NKHVSILC IFAARI YDDQGRQEGRLRHDGAKIAMMNYESFECWDFTVDRQGRFQPWPDLDEHSQALSGLRAI (SEQ ID NO:58)
(underline: nucleic acid editing domain; double underline: cytoplasmic localization signal)
Human APOBEC-3G:
MKPHFRNTVERMYRDTFSYNFYRNPIRLNRTTWLCTEYVKTGPGSPRPPLDAKIFRGOVYS_ELKYHPEMFHWFSS KNRKLHRDGQYEVWTYI_SWSPTKCTRDMATLAEQDVPKTFLIFVARLYYFFDPDYQALRSLQKRDGPRATMK IMNYDEFQHCWSKFVYQRLEPWNMLPYYIIIHLMLGEILRHSMDPFTTFNFNNPWRGRHETYLCEVE RMRHNDTWLNLQRGFLCNPQHPKHLGEHW_AHELCLFLVIPFKLDLDQYRTVCFTSWSPFESC_AQEMAKFIS KNNKVSLCIFITARIYDQGRCQEGLRLTAEGAKI_SIMTYSEFKHCWDTDFVHQPCFQPDGLDESHQDLSGRLRAiLQNQEN (SEQ ID NO:59)
(underline: nucleic acid editing domain; double underline: cytoplasmic localization signal)

Human APOBEC-3F:
MKPHFRNTVERMYRDTFSYNFYRNPIRLNRTTWLCTEYVKTGPGSPRPPLDAKIFRGOVYSQPEH_SAEMCFLSWF CQNQLPAYKCFQITFWFVSWTPCDVAKLAEFLHSPYNWLTIL_SARALYYYRQDRLCRSLQAGARVTIMDY EEFAYCWFVNXENPQEMPWYKFAQHFLRLKEILRYMLDPTTFNFNDPLVLRARRQTYLCEVERLDD GTVLMDQHMGFLCNEAKNLLCQFY_SAHELCFLSWFCDCLSLNPNTKQYWTYSWSPCPCAGEVAFERLARHSNVNT IFTPALLYFFWDTDYQEGLRSLQASGEVIMARKDFYDVCWMDREVNDKFKWGLKYNFLDLSDKQLQEILE (SEQ ID NO:60)
(underline: nucleic acid editing domain)

Human APOBEC-3B:
MNPQIRNPMERMYRDTFSYNFYRNPIRLNRTTWLCTEYVKTGPGSPRPPLDAKIFRGOVYSQTY_EAEMCFLSWF CQNQLPAYKCFQITFWFVSWTPCDVAKLAEFLHSPYNWLTIL_SARALYYYRQDRLCRSLQAGARVTIMDY EEFAYCWFVNXENPQEMPWYKFAQHFLRLKEILRYMLDPTTFNFNDPLVLRARRQTYLCEVERLDD GTVLMDQHMGFLCNEAKNLLCQFY_SAHELCFLSWFCDCLSLNPNTKQYWTYSWSPCPCAGEVAFERLARHSNVNT IFTPALLYFFWDTDYQEGLRSLQASGEVIMARKDFYDVCWMDREVNDKFKWGLKYNFLDLSDKQLQEILE (SEQ ID NO:61)
(underline: nucleic acid editing domain)

Human APOBEC-3C:
MNPQIRNPMKAMYPTGFYQFKNMLWEANDRNETWLCFTVEGIKRRSLLWDTGTVRQFQVYFKQY_SAEMCFLSWF CQNQLPAYKCFQITFWFVSWTPCDVAKLAEFLHSPYNWLTIL_SARALYYYRQDRLCRSLQAGARVTIMDY EEFAYCWFVNXENPQEMPWYKFAQHFLRLKEILRYMLDPTTFNFNDPLVLRARRQTYLCEVERLDD GTVLMDQHMGFLCNEAKNLLCQFY_SAHELCFLSWFCDCLSLNPNTKQYWTYSWSPCPCAGEVAFERLARHSNVNT IFTPALLYFFWDTDYQEGLRSLQASGEVIMARKDFYDVCWMDREVNDKFKWGLKYNFLDLSDKQLQEILE (SEQ ID NO:62)
(underline: nucleic acid editing domain)

Human APOBEC-3A:
MEASPASPRHLMDHIPFNTSFNNINGIRKHKLYCDEVYELDQGDMQHRFGFLNQAANLLCGFYGGR_HAEELRF LDLVPSIQDLLDAQIYVRVTWFISWSPCFSWGCACEVRFALQENTVHRLRIFAIRYDYDPLYKEALQMLRDAQVAQV SIMTYSEFKHCWDTFVHQPCFQPDGLDESHQDLSGRLRAiLQNQEN (SEQ ID NO:63)
(underline: nucleic acid editing domain)

Human APOBEC-3H:
MALLTAETFRLQFNNKRLRRPYYPKALLCYQLTPQNGSTPTGRYFENKKC_HAEICFNEIKSMGLDETOCYQ VTCYLTWSPCSACWELDFIKAHDLINLGIFASRLYHHWCKFKQQKGLALLCSQVPVEVMGFKFADCWMDVFV HKEKPLSFNPYKMLEELDKNRSAIKLRLEIKIPGPVRAQGRYMDILAECV (SEQ ID NO:64)
(underline: nucleic acid editing domain)
Human APOBEC-3D:
MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGPVLPKRQSNHRQEVYFRF
ENHAEMCFLSFCGNRLPANRFQITQNFVSWNPCPCVVKVTKFLAEHPNVTILIAARALYRVDRODWRWLRLHLKAGARVKM
YDREFAYCWENFVQACQFQMPWYKFDDNYASLLHRTKEILRNPEAMYPHIPFYFHKNNLKACG
RNEWLCFTMEVYKBSMKDGKFRGVRDNQVFTHC_HAEBCFLSFCDILSPTNHYETWYSWSPCPCAGEV
AEFLARHSVNLITFTRALCFCWDTDYQEGCLSLSQEGASVKIMGYKDFVSCWNFVYSDDEPFPKWGLQTNRFFLLKRLREILQ
(SEQ ID NO:65)
(underline: nucleic acid editing domain)

Human APOBEC-1:
MTSEKGPSTGDPTLRRIEPWEFDFVFYDPREDLRKAECLLYYEIKWGMRSKRWGSGKNTTNSHEVFNFKFTSERDF
FHSMSCSITWFLSWSPCGCESAITEFSLHRPHYTLVIYVARLFWHMDQONRQGRLDVNSVGTIQIMRASEYY
HCWRRNFVNPQDEAHWPQYVPLWMMLAEALHCILSSLPPCLKIKSRWQNHETFRLFRHLQNCYIQIPHIPHLA
TGHLPSVAAWR (SEQ ID NO:66)

Mouse APOBEC-1:
MSSETGPVAVDPTLRRIEPHEFENVDFPDRLRKECLLYEIKWGMRSKRWGSGKNTTMHEVFNFKFTTERTY
FRCPRSTCITWFLSWSPCGCESAITEFSLHRPHYTLVIYVARLFWHMDQONRQGRLDVNSVGTIQIMRASEYY
HCWRRNFVNPQDEAHWPQYVPLWMMLAEALHCILSSLPPCLKIKSRWQNHETFRLFRHLQNCYIQIPHIPHLA
TGHLPSVAAWR (SEQ ID NO:66)

Mouse APOBEC-1:
MSSETGPVAVDPTLRRIEPHEFENVDFPDRLRKECLLYEIKWGMRSKRWGSGKNTTMHEVFNFKFTTERTY
FRCPRSTCITWFLSWSPCGCESAITEFSLHRPHYTLVIYVARLFWHMDQONRQGRLDVNSVGTIQIMRASEYY
HCWRRNFVNPQDEAHWPQYVPLWMMLAEALHCILSSLPPCLKIKSRWQNHETFRLFRHLQNCYIQIPHIPHLA
TGHLPSVAAWR (SEQ ID NO:66)

Rat APOBEC-1:
MSSETGPVAVDPTLRRIEPHEFENVDFPDRLRKECLLYEIKWGMRSKRWGSGKNTTMHEVFNFKFTTERTY
FRCPRSTCITWFLSWSPCGCESAITEFSLHRPHYTLVIYVARLFWHMDQONRQGRLDVNSVGTIQIMRASEYY
HCWRRNFVNPQDEAHWPQYVPLWMMLAEALHCILSSLPPCLKIKSRWQNHETFRLFRHLQNCYIQIPHIPHLA
TGHLPSVAAWR (SEQ ID NO:66)

Human ADAT-2:
MEAAKAPKGAAGCSVSAAETEKMWEAMHMEAKEALTEVFPGCLMVYNENVKGKRNEVQTNAIRHAEV
AIDQVELDRCQSKSPSEFVHTLYVTVEPICMAAALRMLKIPLVYGQNERFGGCSVLNISADLPNTRG
PFQCIPGYPRAEEMAVLKFTYKYQENPNAPSXKVKKECQK (SEQ ID NO:69)

Mouse ADAT-2:
MEEKVESTTPDPCVSVQETEKMWEAMHMEAKEALTEVFPGCLMVYNENVKGKRNEVQTNAIRHAEV
AIDQVELDRCQSKSPSEFVHTLYVTVEPICMAAALRMLKIPLVYGQNERFGGCSVLNISADLPNTRG
PFQCIPGYPRAEEMAVLKFTYKYQENPNAPSXKVKKECQK (SEQ ID NO:70)

Mouse ADAT-1:
MWTADIEAQLCAHYNVRLPKQGKPEPNNREWTLAAAQKIQASANQACDIEKEFQVQVTEVSMGTCGTICQSK
MRESGIDLNSHAEIIRASSFRQRYLHLHLALAVKLESIFVPGTQGRLWRLRFDLSFVFFSHTCCGADSSIIPM
LEFEQEQQCPPISWANNSPVQETENLDSKDRNCEPASPVAKMRLGTPARLSNCVAVHTGQESGPKFVDV
SSDDLTKEEPDAAANGASGSRFVVDYTGAKVCPGETDLREPAGAYQVGLRLVVKPGGRDRCMSSCCSMKMNAR
MWVQLCPQALLMMHFLKPIYLSAVIGKCPYSEQAEAMRALTGRCEETLVLFPGFQVQLEIQSGLFQSCAV
HRRKGRDSGPRCLPVCGAASIWAWPQQPLODLVTANGFPQGTTEiguISPRSRKLIWTKKDEQPD
SIRVTKKLDTQWEYKDAASYEQAEWALRRTPFASwirNPDDYHOFK (SEQ ID NO:71) (underline: nucleic acid editing domain)
Human ADAT-1:

MTADEIAQLCYEHGIRLPKKGPEPNHEWTLAAVVKIQSPADKACDTPDVPQVTKEVV SMTGTCICIGQSKMRKNGDI LNDSHAELAARRSFQRYLLHQLQLAATLIKEDSIFVPGTOGKVWRDL IEFVFS SHTPCGDAS1IPM LEPEDQCPVFRNWAHNS SVEAS SNLEAPNKRCEPDPSVTKKMRLEPGTAARVEYINGAHHQSGFKQKSPG I SPGIIHCSTLTVETALVTRIAPGSAKVI DVEIRTGAKCVPGEAGDSGKPAHQLQRVVKPGDRTRSMCS DKKMARNVNLGCQALLMHILLEPI YLSAVVIGKCPYQESAMQRAL IGRCQNYSAKPFGVQELKI_LQSDLLFEO

[underline: nucleic acid editing domain]

[0092] In some embodiments, the enzymatic domain comprises one or more of a transcriptional activator. For example, in some embodiments, the general architecture of exemplary dCas9 fusion proteins with a transcriptional activator domain comprises the structure:

\[
[NH_2]-[\text{NLS}]-[\text{Cas9}]-([\text{transcriptional activator}]_n)-[\text{COOH}],
\]

\[
[NH_2]-[\text{NLS}]-([\text{transcriptional activator}]_n)-[\text{Cas9}]-[\text{COOH}],
\]

\[
[NH_2]-[\text{Cas9}]-([\text{transcriptional activator}]_n)-[\text{COOH}],
\]

\[
[NH_2]-([\text{transcriptional activator}]_n)-[\text{Cas9}]-[\text{COOH}];
\]

wherein NLS is a nuclear localization signal, dCas9 comprises an intein as provided herein, NH₂ is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein. In some embodiments, the fusion proteins comprises one or more repeats of the transcriptional activator, for example wherein \( n = 1-10 \) (e.g., \( n \) is 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10). In some embodiments, \( n = 1-20 \). In some embodiments, a linker is inserted between the dCas9 and the transcriptional activator domain. In some embodiments, a linker is inserted between the NLS and the transcriptional activator and/or dCas9 domain. In some embodiments, the NLS is located C-terminal of the transcriptional activator and/or the dCas9 domain.

Additional features, such as sequence tags, may also be present. In some embodiments, the transcriptional activator is selected from the group consisting of VP64, (SEQ ID NO:73), VP16 (SEQ ID NO:74), and p65 (SEQ ID NO:75).

VP64
GSGRADALDFDLDMLGS DALDDFDDLMLGSDALDDFDDLMLGS DALDDFDDLML1N (SEQ ID NO:73)

VP16
APPTDVSLGDELHLDGEDVAMAHADALDDFDLDMLGDGDSPGPGFTPHDSAPYGALDMADFEFEQMFT DALGI DEYGGEFPGR (SEQ ID NO:74)
[0093] In some embodiments, the enzymatic domain comprises one or more of a transcriptional repressor domain. For example, in some embodiments, the general architecture of exemplary dCas9 fusion proteins with a transcriptional repressor domain comprises the structure:

\[
[NH_2] - [NLS] - [Cas9] - [(transcriptional repressor),] - [COOH],
\]
\[
[NH_2] - [NLS] - (transcriptional repressor), - [Cas9] - [COOH],
\]
\[
[NH_2] - [Cas9] - (transcriptional repressor), - [COOH], or
\]
\[
[NIL] - (transcriptional repressor), - [Cas9] - [COOH];
\]

wherein NLS is a nuclear localization signal, dCas9 comprises an intein as provided herein, NH2 is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein. In some embodiments, the fusion proteins comprises one or more repeats of the transcriptional repressor, for example wherein \( n = 1-10 \) (e.g., \( n = 1, 2, 3, 4, 5, 6, 7, 8, 9, \) or 10). In some embodiments, \( n = 1-20 \). In some embodiments, a linker is inserted between the dCas9 and the transcriptional repressor domain. In some embodiments, a linker is inserted between the NLS and the transcriptional repressor and/or dCas9 domain. In some embodiments, the NLS is located C-terminal of the transcriptional repressor and/or the dCas9 domain. In some embodiments, the NLS is located between the transcriptional repressor domain and the dCas9 domain. Additional features, such as sequence tags, may also be present. In some embodiments, the transcriptional repressor is selected from the group consisting of the KRAB (Kriippel associated box) domain of Koxl, SID (mSin3 interaction domain), the CS (Chromo Shadow) domain of HP1a, or the WRPW domain of Hesl. These and other repressor domains are known in the art, and in some embodiments correspond to those described in Urrutia, KRAB-containing zinc-finger repressor proteins. *Genome Biol.* 2003;4(10):231; Gilbert et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell.* 2013; 154, 442-451; Konermann et al., Optical control of mammalian endogenous transcription and epigenetic states. *Nature.* 2013; 500, 472-476; and published U.S. patent application U.S.S.N. 14/105,017, published as U.S. 2014/0186958 Al, the entire contents of which are incorporated herein by reference. In some embodiments, the transcription repressor domain comprises one or more repeats (e.g., 2, 3, 4, 5, 6, 7, 8, 9, or 10
repeats) of a KRAB domain. In some embodiments, the KRAB domain comprises an amino acid sequence selected from the group consisting of SEQ ID NOs:76-79. In some embodiments, the transcriptional repressor domains comprises one or more repeats of a SID protein. In some embodiments, the SID protein comprises an amino acid sequence set forth as SEQ ID NO:80. In some embodiments, the repressor domain comprises 2, 3, 4, 5, 6, 7, 8, 9, or 10 repeats of a SID protein (e.g., SEQ ID NO:80). In some embodiments, the repressor domain comprises four repeats of SID (e.g., SID4x; SEQ ID NO:81).

**KRAB** (human; GenBank: AAD20972.1)

MNMFKEAVTFKDVAFAFEEELGLLGPAPQRKLYRDVMVENFRNLLSVGHPFPKQDVSPIERNEQLWIMTTATRRQGNLDTLPVKALLOWLYDLAQ (SEQ ID NO:76)

**KRAB** protein domain, partial (human; GenBank: CAB52478.1):

EQVSFKDVCVDFTQEEYWLLDAQQKI LYRDVI LENYSNVLSVGCITKPEVI FKIEQGEEPWILEKGFPSQCHP (SEQ ID NO:77)

**KRAB** A domain, partial (human; GenBank: AAB03530.1):

EAVTFKDVAVFTEEELDPAQRKLYRDVMLENFRNLLSV (SEQ ID NO:78)

**KRAB** (mouse; C2H2 type domain containing protein; GenBank: CAM27971.1):

MDLVTRYDDHVNFTQDEWALLDSPQKSYKGMLETYKLTAIGYIWEHT1 EDHFQTSRSHGNKKT (SEQ ID NO:79)

**SID** repressor domain:

GSGMN1 QMLLEAADYLEREAEHGYASMLP (SEQ ID NO:80)

**SID4x** repressor domain:

GSGMN1 QMLLEAADYLEREAEHGYASMLPGSGMNIQMLLEAADYLEREAEHGYASMLPGSGMNIQMLLEAADYLEREAEHGYASMLPSR (SEQ ID NO:81)

[0094] In some embodiments, the enzymatic domain comprises an epigenetic modifier or a catalytic domain thereof. For example, in some embodiments, the general architecture of exemplary dCas9 fusion proteins with an epigenetic modifier or domain comprises the structure:

[\text{NH}_2]-[\text{NLS}]-[\text{Cas9}]-[\text{epigenetic modifier}]-[\text{COOH}].
[NH2]-[NLS]-[epigenetic modifier]-[Cas9]-[COOH],
[NH2]-[Cas9]-[epigenetic modifier]-[COOH], or
[NH2]-[epigenetic modifier]-[Cas9]-[COOH];

wherein NLS is a nuclear localization signal, dCas9 comprises an intein as provided herein, NH2 is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein. In some embodiments, a linker is inserted between the dCas9 and the epigenetic modifier domain. In some embodiments, a linker is inserted between the NLS and the epigenetic modifier and/or dCas9 domain. In some embodiments, the NLS is located C-terminal of the epigenetic modifier and/or the dCas9 domain. In some embodiments, the NLS is located between the epigenetic modifier domain and the dCas9 domain. Additional features, such as sequence tags, may also be present. Epigenetic modifiers are well known in the art, and typically catalyze DNA methylation (and demethylation) or histone modifications (e.g., histone methylation/demethylation, acetylation/deacetylation, ubiquitylation, phosphorylation, sumoylation, etc.). The presence of one more epigenetic modifications can affect the transcriptional activity of one or more genes, for example turning genes from an "on" state to an "off" state, and vice versa. Epigenetic modifiers include, but are not limited to, histone demethylase, histone methyltransferase, hydroxylase, histone deacetylase, and histone acetyltransferase. Exemplary epigenetic modifying proteins can be found in Konermann et al., Optical control of mammalian endogenous transcription and epigenetic states. Nature. 2013; 500, 472-476; Mendenhall et al., Locus-specific editing of histone modifications at endogenous enhancers. Nat. Biotechnol. 2013; 31, 1133-1136; and Maeder et al., Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. Nat. Biotechnol. 2013; 31, 1137-1142; the entire contents of each are incorporated herein by reference. In some embodiments, the epigenetic modifier domain is LSD1 (Lysine (K)-specific demethylase 1A) histone demethylase, which in some embodiments, comprises in whole or in part, an amino acid sequence set forth as SEQ ID NO:82 or SEQ ID NO:83. In some embodiments, the epigenetic modifier domain is TET1 hydroxylase catalytic domain, which in some embodiments, comprises an amino acid sequence set forth as SEQ ID NO:84. In some embodiments, the epigenetic modifier is a histone deacetylase (HDAC) effector domain. In some embodiments, the HDAC effector domain comprises in whole in part, an amino acid sequence corresponding to any of the HDAC effector proteins provided in Supplementary Table 2 of Konermann et al., Optical control of mammalian endogenous transcription and epigenetic states. Nature. 2013; 500, 472-476; SEQ ID NOs:85-96. In some embodiments, the epigenetic modifier is a histone
methyltransferase (HMT) effector domain. In some embodiments, the HMT effector domain comprises in whole in part, a amino acid sequence corresponding to any of the HDAC effector proteins provided in Supplementary Table 3 of Konermann et al., Optical control of mammalian endogenous transcription and epigenetic states. Nature. 2013; 500, 472-476; SEQ ID NOs:97-106.

LSD1, isoform a (human):
MLSGKKAAAAAAAAAAAAATGTEAGPTAGGSENGSEVAAQPAQLG SGPAEVPGAVGERTPRKKEPPRA SPPGGLAEPPSGAPQAGPTVPGSATPMETGIAETPGRTSRRKRKVEYREMDESLEANLSEDEYY SSEEERNAKAEKKLPPPPPQAPPPPEEE SEPEEPSEQAGGLQDDSGSGYGDGQAGSVEGAAAFQSLP HDRMTSQEACFPD I SGPPQQTQVKLF I RNRTLQLWLDPKQLTQETALTQLEAPYNSDTVLVHHRV HSYYLERHFL INFIC YKRI KPLPTKGTGV I 11 GSYSGLAARQLQSGFMVDMTLEARDVRGVRATF RKGNYVADLGMWTLGNGPMASQKVNMLAKIKQCKPLYEANGQADTVKVPKEKDEMEQEENR LLEATSYLQLDFNVLNKPVSQALEWILQKHEQKVKDEI EHHWKKIVKTQEEKELNNKMNLK EKIKELHQQYKEASEVKKPR I 1 AEFVLVSKHRDLTLALCKEYDELATQGKLEELQLEELANPSDVY LS SRDQ I LDWHFANFANATPL TSLKHWDQDDDEFTGSHLTEVRNGY S CVPVALAEGLI KLNT ARVQVRVYTAS GCEVIAVNTRS TSQTF IYKCDAVCLTPLGLKQQPAPQFVPLPEWKTSAVQMRGF GNKNKVVLCFDRVFWDPSVNLFGVHGSTTASRGELFLFWNLKAPI LLLAVAGEAAGIMENI SDDIVV GRCLAI LKGI FGS SAVQPQPKETWSRWRADVPSYVAAGS SGNDYDLMAQPI TPGPS IPGAQPQ IPRLFFAGEHT IRNYPATVHGALLS GLREAGRIADQFLGAMYTLPRQATPGVPAQQPSM (SEQ ID NO:X)

LSD1, isoform b (human):
MLSGKKAAAAAAAAAAAAATGTEAGPTAGGSENGSEVAAQPAQLG SGPAEVPGAVGERTPRKKEPPRA SPPGGLAEPPSGAPQAGPTVPGSATPMETGIAETPGRTSRRKRKVEYREMDESLEANLSEDEYY SSEEERNAKAEKKLPPPPPQAPPPPEEE SEPEEPSEQAGGLQDDSGSGYGDGQAGSVEGAAAFQSLP HDRMTSQEACFPD I SGPPQQTQVKLF I RNRTLQLWLDPKQLTQETALTQLEAPYNSDTVLVHHRV HSYYLERHFL INFIC YKRI KPLPTKGTGV I 11 GSYSGLAARQLQSGFMVDMTLEARDVRGVRATF RKGNYVADLGMWTLGNGPMASQKVNMLAKIKQCKPLYEANGQADTVKVPKEKDEMEQEENR LLEATSYLQLDFNVLNKPVSQALEWILQKHEQKVKDEI EHHWKKIVKTQEEKELNNKMNLK EKIKELHQQYKEASEVKKPR I 1 AEFVLVSKHRDLTLALCKEYDELATQGKLEELQLEELANPSDVY LS SRDQ I LDWHFANFANATPL TSLKHWDQDDDEFTGSHLTEVRNGY S CVPVALAEGLI KLNT ARVQVRVYTAS GCEVIAVNTRS TSQTF IYKCDAVCLTPLGLKQQPAPQFVPLPEWKTSAVQMRGF GNKNKVVLCFDRVFWDPSVNLFGVHGSTTASRGELFLFWNLKAPI LLLAVAGEAAGIMENI SDDIVV GRCLAI LKGI FGS SAVQPQPKETWSRWRADVPSYVAAGS SGNDYDLMAQPI TPGPS IPGAQPQ IPRLFFAGEHT IRNYPATVHGALLS GLREAGRIADQFLGAMYTLPRQATPGVPAQQPSM (SEQ ID NO:X)

TET1 catalytic domain: SIVQVLSRDPDPAAPALTNDHLVAACLGLGRPALDDAVKKGLPHAPAI IKRTNRPERTSHRVDADQVYRVLGFQQCHSHPAQADFDMATIQFGMSGGGSHLPSCCLDRLVQDKGYPYTHLAGPSVAARVE EN RYGQKGNARIA IVYTQGKCEGSG SDEEKLVLQRGHTCPTAVMWL IMWVD GIPLMAMDRTELTENKLNYNGHPDTRRCLNEERTCICGI DPECTGASFSFGC SWMYFNGCKEG RSPPSRPFR I DPS SPLHEKNLDDLQLASLATHPIYQYAPVAYQQNYEVENARECGLSKGEPF S GVTAICLDLCAHPRHRDHNMGNSTTVCTLTREDNRLGVI PQDEQHLVPLYKLDSTDGESEKGMEM A KIKSGAEVFALRTKRIKSKRTCTFQTPVRSPKGGKRAAMMETVLAIHKAVEKKPI PRIKRKNSTTTTNSKP S LSLPTLSNTEVTQPEVKSETEPH I 1 LSKSNDNTYES IMPSAPHPVKEASGFSWPSKATAPLKNDATASCGFSERS STPHTCTMPSGR SGANAAADGPGI SQLGEVAPLTPLSAPVMEPL INSEPSTVE EPLTYPHQPSFLSPQDLAS SMPEDEQHSEADEEPS DEPL SDDPLPSAEKLP DEYWSDESE I FLDAO GVGVAIAAPAHGSL I ECARRHELATTPVHEPNRHNHPRLS LSFQYHKNLNKOPQHGFELNKL K
FEAKEAKNKKMKASEQKDQAANEGPEQSSEVNELNQIPSHKALTLTHDNWTSPYALTGVAGPYNHW
v (SEQ ID NO:X)

HP AC effector domains:

HDAC8 (X. laevis):
ASSPKKKRKVEASMERSWPKVASMEEMAALFHDAYLQHHLKVEEGDNDDPETLEYGLGYDCITEG
IYDAAAAGAVGLTIAAEQLIEKTRIAVNWPPGWVHAKDEAGSGFCYNDAVLGIKLREKFDRLVYV
DMDLHHDGVEDAFTSVKMTSLKSFIPEGFFGFTQGDSDIGLQKRYSSINPIQWQDGQGKYYQI
CEGVLKEVFPTFNEAPWLGADITIAAGDMCSFNMTPFEGIGKCLLYQWQPLTL1LGGGYHLFPNT
ARCWYLTALIVGRTLSSEIPDHEFFTEYPDVYLETIFPCRDRNDTQVKQEILS1KGNLKRWEF
(SEQ ID NO:85)

RPD3 (S. cerevisiae):
ASSPKKKRKVEASRVAYFAYVADGVNYAYAGHFMKPHHRMAHSILMNYYLGKKEIRAKPQ
MCQFHTDEVYIDFLSRTPDONLMEFKRESVKNVNGDCCPVDGLEYECYSSGSMGAEALNRGKD
AVNYAGGLHAKKESAEGCUPYIDLOIVHVGEEVAYFTTDRWMTCSFSH
KYGEEEFGTIGELRDIGVAGKNGYAVNVLPLRDGIDDAYTSVFEPVIMKIMEWYQPSAWLQC
GDSLGSGLCFNLSMEGHANCNYKSFFQFPMMWGGGYTMMRNVARTWETFGLNNLKDLPYEF
(SEQ ID NO:86)

MesoLo4 (M. loti):
ASSPKKKRKVEASMLQIVIPHFYDAGFATNHRFPMKSYPLMLAMALRAGLSLDALTITEFAASWL
KLHAADAVYGDVSCEPVEKIIIFGVPVRLAQRATGTLIAALARLHIGACNTAGSHHAR
RAQAGAIFCTFNDVAVASLVLLEDGAQQNLWNLDDLVHQGDGTADILSDEPGFTSFNSMEHHYRP
KASIADLIALDGFDAAYLRLALJLFELSARARWIVYAGVHAEDRLGLALGNSGLARYRDEMV
IGHFRALGIPCQVGGYSTDVFALASRHAILEFVASTYAEF (SEQ ID NO:87)

HDAC11 (human):
ASSPKKKRKVEASMFLHTQLYQHPETRWPVIVSPRYNIFFMGEKLHPDAGKWFKVINFLKEKL
LSDSMVEAREASEDLLVHHTRYNLKSWFPAVATITEIPVIFLPNFQKRKLVPRLRTQTTGTM
AGKLAEVRGWAINVGHHCCSDRGGFCAYADITLAIKFLPVEGGRATI IDDLAQPQNGHERD
FMDDKRVYIMDVYNRHPGDFRAFQAKAIRKKEWTEGDDEYDLKVERNIIKSLQEHLPMDVYAG
TDILEGDRGLGLISI SPAGIVKRDDELVFRVMGRVPMILMTSSGGYQKRTRI IADS ILNLFGLLGIGP
ESPVSVAQNSDTPLLPAPVPEF (SEQ ID NO:88)

HDT1 (A. thaliana):
ASSPKKKRKVEASMEFGVEVKSGPKVTVEPEEGLIHIQVQSLGEGNNKKGGEFMIPHLVKVGNQNLVL
GLTSTENIFQIFDLDVFDEKFELSTWGGKSVYFVQKTPQIENPQGYSEEEEEEVEEFAGNAAKAVA
KPKAFAKVEPCKAVDEESDSDGMDDEDSGDSEEEDEEKPFKKAFSSEKRENTTPKAPAVSAKAK
AVTPOKTDDEKKGGKAANQSEF (SEQ ID NO:89)

SIRT3 (human):
ASSPKKKRKVEASMGAGIPTSIGIDPSRFSPSGLYSLONLQYDLPYEAIFELLFPFFHNPKFFTLA
KELYPNGKPNVTHYFLRIIHLKLILRLYRTQHDQULEDVSSGAPKLVFEAHGFASATCTVQRPFF
GEDIRADVMADRFPSCPVFCTGVDJIVFPGEMLQFQRLWHDDFMADLILLILGTSVLEFPEASLTEA
VRSSVFRLLINRDLGFLAWHPFRSDVAQLGVDVHGVESLVELLLGCTEMRDLVQRETGKLGDPIKF
(SEQ ID NO:90)
HST2 (S. cerevisiae):
ASSPKKKRKVEASTEMSVRKIAAHMKSNPNAKVIFMVGAGISTSCGIPDFRSPGTGLYHNLARLKLPY
PEADEVDFDFQSDPLFETYLAKELYPFGNFRSKFYHLLKLQDFKDLRKYTVQNI9IDLERQAGVKDDL
IIEAHGSAHCHICCGKGYYPQVSKSLAEHFIFIKDFVCKDCGELVKPAIFVGEFLDFSDSETWLN
DSEWLRKETTSKGPQQPQLVWTVELAVPSAPLIEEPRKVRVLNLTVGDFKANKRPTDIL IV
HQSDEFAEQIEVLEEGWEDFQEDKILTAQGMGEF (SEQ ID NO:91)

CobB (E. coli (K12)):
ASSPKKKRKVEASMEKPRVLVLTGAGISAESGIRTFRAADGLWEEHRVEDVATPEGFDRDPELVQA
FYYARRQLOQPEIQPNAAHLALAKQLQDALGDRFLLLVTQNI9IDLHERAGNTVI9HMHGLKLKVR
CSQSGQ VLDWTGDPEDKHCQCPFAPLRHWWGEMPLGMDEIYMAISDFIAIGTSGHVPAAGFVHE
AKLHGAAHTVELNLEPSQVNFEAEKYGPASQVQVFEPVEKLLKLKAGSIAEF (SEQ ID NO:92)

HST2 (C. albicans):
ASSPKKKRKVEASMPSLDDILKPGAEEAVKNGKVTTFNGAGISTGAGIPDFRSPDGLYANLAKNLPL
FEAEYDIDFFKYPFFYPFGNYPFQFHI9KLQDQGSLKRVPYTNQI9IDLTERLAGVDEK
YVHEAHGFASNNHCVCHKEMTETLKTMDKCPHSCQGCVGYKPDIVFFGELPGKFFDLWEDDC
EDVEVAIGATSLLTFPPASLPPGEVNKNKCLRLVNEKVGTFKHEPRKSL IALHDCDVAERLC
GLDDKLNEYKEEIKYSKAEIKMEIIEDKLKEEHLKEDKHTTKVDKEKQNDANDKEEQLID
KAAEF (SEQ ID NO:93)

SIRT5 (human):
ASSPKKKRKVEASSSSMADFRKFFAKAHIVI ISAGAVSAESGVPTFRGAGGYWRKWQAQLATPLAF
AHNPSRVRWFYHYREVMGSKEPNAGHRAIAECETRLGKQGRWITQNI9IDLHERAGNTVI9HMHGL
KSQSGQ VLDWTGDPEDKHCQCPFAPLRHWWGEMPLGMDEIYMAISDFIAIGTSGHVPAAGFVHE
AKLHGAAHTVELNLEPSQVNFEAEKYGPASQVQVFEPVEKLLKLKAGSIAEF (SEQ ID NO:93)

Sir2A (P. falciparum):
ASSPKKKRKVEASMGNLMISFLKKDTQSITLEELAKI QKKCKHVVALTGSSTAESNIPSFRGSSNS I
WSKYDPRYGIWGWKYPKEWIRVDIRI SDYIEINNNGHVALSTLESGLYKSVTQNVDGLHVRSG
GNTKVI SLHGNVFEEAVCTCTNLKVINKLMLQKTSFHMLLEPFCGCGIPKNI ILEGEWSSDLDK
EAEEIIKACDLLLVLTVGTSSTSTVSTANDLFACACKKKKVEINISKTYITNKMDSYHVCAKSETLKVA
NILKGS SEKNNKIMEF (SEQ ID NO:94)

SIRT6 (human):
ASSPKKKRKVEASMSVNYAAGLSPYADKGKCGLPEDPPEELERKVELRLWQSSSFWHTGAGI
STASGIDPGFRGPHGWMTMEGRLAFLKDPFDTEFESRPQTOMALVQELRGFLVQNSVQDHLHVSQ
FPFROKLALHGNMFVEECACKCTQYRDVTGWMKATGRCLTVAKARGLRACGRELRTILDWE
DLPRDRLALADEASRNALDSITLQSIQRPSGNFLATKRRGRLVNLQPTKHDRAHLDIRHYV
EDVMTRLMKHLGIEPAPDGRVLERALPPEF (SEQ ID NO:96)

HMT effector domains:
NUE (C. trachomatis):
ASSPKKKRKVEANMTNSTDLYLSHGGDSAIPFVPPRQVLEQLQQFSLPELQFQNAAVKQRIQL
CYYEKKLAVAISLLKLQHLXHQRRAKPNFVAICWINSYVGYGFVES IAPWSYIGETGLRRR
QALWLDENYCFRYFPYRSPFRYFTIDSQGMQGNTFRFINHSDNPNLEAIAGFENGIFHI IIRAIKDIL
PGEELCYHYGPLWKRKREEFVPQEEF (SEQ ID NO:97)
vSET (*P. bursaria chlorella virus*):

```
ASSPKKKRKVEASMFNDVRVIKVKSPLGGYGVFARKSFEGELVEECICVRHNDWDGTALEDYLF
R NMSAMALGFGAIFNHSDKPANHELTAALGKRMRFITKPIAIGEEIT SYGDDYWLSRPRLTQNEF
(SEQ ID NO:98)
```

SUV39H1 (human):

```
ASSPKKKRKVEASNLKCVRILKQFHKDLERELLRRHRSKTPHRHLDSLANYLVQAKQRAALRRWEQ
ENAKRHSLHGRITVENEVDLGDGPPFRAFYVINEYRVGEGIQLNQAVGCEQCDLAWPTGCGCCFGASLH
KFAYNQDGVQRVLRAGLPI YECSNRRCGYDCPNWRQWKIRYDLCIFTRDDGRGWGRVTRLEKIRKNSF
VMFYVGEI ITSEEAEARRQYIQDRGQATYLFIDLDYVEDVYTVADAAAYGNISHFVNHSCDPLNLQVNYVF
IDNLERLPRIAFFATRATIRAGEELTDFYNMQVDPVDMESTRMDNSNFGLAGLPGSPKKRVRIECKGTE
SCRKYLF E (SEQ ID NO:99)
```

DIM5 (*N. crassa*):

```
ASSPKKKRKVEASMEKAFRPFFHHGKPDANPEKNNCHWQCQIRSFATHAQLPI ISVNRREDDAFLNP
FFRDHSI IGKNVVPADSFVRGVCASDDECYMYSTCQLDEMAPDSEEADPYYTRKKRFAYYSQAK
KLGLRDRVLQSQEFPEYEHCQGACSCPDCNPWNERGRTVPLQIFTRKDGRGWVCPVNIKRGQFVDRY
LGEI ITSEEdRRRAESTIARRKDYFALDFSDPSLDPMLAQIPFLEVDGMYSMGPFTRHENCEDP
MNAIFARVGDHADKHDILAFALKDIPKGETLTDFYNQLTGESDAHDPKRIEMTKCLCTAKCR
GYLWFE (SEQ ID NO:100)
```

KYP (*A. thaliana*):

```
ASSPKKKRKVEASDISGGLEFKGIPATNRVDVSPVSTSGFTYIKSL IEPNI IPKSTSTGCNCRGSC
TDSKCAKCLNLGGNFPYDVNLNGRLIESRDFWECGPHCGCPKCVNRTSQKRLRFNFLVEFRSAKKG
WAVRSEWYIPAGSPVCEYIMGVRATDVIISDEYIFIDECDQTMQLGGRQRRLVPAHMPNNGVS
QSSEDENAPFIDAGSTGNFARHENSEEPFLTFCQVLSSQDRQLARLFAADNISPMQELTYDY
GYALDSVFE (SEQ ID NO:101)
```

SUVR4 (*A. thaliana*):

```
ASSPKKKRKVEASQASYLHVLSLARI SDEDCANCKGNCLSDADFPCTCARETSGYAYTKEGLLKEKFL
DTCGLMKEPSFKVQKDCPLERHDHYGKYRDCGLHILRKFIKECWRCGDCMQGNCNWRQGRGICR
QLQYVFTQEGKGNLRLQDLPKPGFIICYEISELMTEDYAVRSRERHPTVPTLDAWDQEGKDL
KDEEALCICACICNVRARFIHANCEDANMIPIEIETPDHRHYIHFITL ERDVKAMECDTLWYMIDF
NDKSHVKAFRCcccGSECRDKIKGSQKISERRKIQVSAKKGQSKEVSKKRKEF (SEQ ID NO:102)
```

Set4 (*C. elegans*):

```
ASSPKKKRKVEASQMQLHEQANISVFTINDIPRSHDSTMPTESLYFDFATTDLDSVLANTFTHMSSK
RRYLYQDEYRARTVMKTFFQRQDWNTA YGLLTLRSVSHFLSLKLPNKLFEFREDHIVRFNLMFILDS
GYTIQECRKYQGEHQQGKLVSTGVNSRDKIERLSGWCLLHSEDSSILAQEGDSFSVMSTRK
STLWLGPGYAIINNCRTPCEFVSHTAHRVLDMVPGEITCFYEGSEFFGPNIDCECTCETKMN
GAFSYLRGNGENAPRIESKKTKYELRSRSEF (SEQ ID NO:103)
```

Setl (*C. elegans*):

```
ASSPKKKRKVEASMVKAAKLATSRMRKDRAAAASPSDDIENSEPSSLASHSSSSHGRMTPSNTRRSR
KGVSVKDNHKTIEFQFPVRNSRNTSKQISDEEAKHALRDFTVLKGTNRELLELYKDWRGIRKVN
FEKGDFWEYRGVMMESEAKVIEEQYNSDEEEIGSMYMMFEHNKKWICDASETSPFWGRHLINHSLVR
PNLKTKWEIDGSHHLLILVARQIAQGEELLYDGDRAETIAKNFWLVNTF (SEQ ID NO:104)
```
SETD8 (human)
ASSPKKKRVEASSCDSTNAIAKQLLKKPIKGGAPRKKAQKTQNRLTDFYVRRRSSRKSIAELQSEERKIDELIESGREEMKIDLDGKRGVATKPSGDFVVEYHDLIEITDAKREALEYQDPSTGCYMYFQYLSTKTVDATRETNRGLH5KCNQTKLHDIDGVPVLILIASRDIAAGEELLY
DYGDRKASIEAFPWLKHEF (SEQ ID NO:105)

TgSET8 (G. gondii):
ASSPKKKRVEASSRRTGEFLRDAQAPSRWLKRSKGTQDDGAFCLETTLAGAGDDAAGGERGRDREGAADKAKQGREERQKELEEREFEEMKEFEEKAQRMIAARRAALTGEISDGGSKSRRVPSLPLNEDDAILIEIIIDPEQGILKWLPSVMSISIRQRTVYQECLELRDLTACILHTKVPGKRATVFAADILKDFFWEYK
GELCEREAREEREQRYNRSKVMGSFMYFKNGSRMMAIDATDKEQDFG4FARLINSRNPMTFRAI
TLGDFNSEPRLIFVARRNIEKGEELLVDGERDQPDVKEHPWLNEF (SEQ ID NO: 106)

[0095] In some embodiments, ligand-dependent Cas9-intein variants are provided herein that exhibit decreased off-target activity. For example, in some embodiment, Cas9-intein variants are provided herein that comprise a Cas9 nuclease domain, or a nuclease-deficient Cas9 domain and a heterologous nucleic acid-editing domain, such as, for example, a heterologous nuclease domain, a recombinase domain, or a deaminase domain. In some such embodiments, the ligand-dependent Cas9-inteins provided herein exhibit decreased, minimal, or no off-target activity in the presence of a ligand at a concentration effective to effect excision of the intein from the Cas9-intein variant, or at a concentration effective to induce a desired modification (e.g., cleavage, nicking, recombination, or deamination) of a target site. In some embodiments, the ligand-dependent Cas9-intein variants provided herein exhibit an off-target activity in their active state (e.g., in the presence of or after being contacted with a suitable ligand) that is decreased as compared to the off-target activity of wild-type Cas9. For example, in some embodiments, the off-target activity of a Cas9-intein variant is decreased to less than 80%, less than 75%, less than 50%, less than 45%, less than 35%, less than 30%, less than 25%, less than 24%, less than 23%, less than 22%, less than 21%, less than 20, less than 19%, less than 18%, less than 17%, less than 16%, less than 15%, less than 14%, less than 13%, less than 12%, less than 11%, less than 10%, less than 9%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, or less than 1% of wild-type Cas9 under the same conditions.

Pharmaceutical compositions

[0096] In some embodiments, any of the ligand-dependent site-specific enzymes described herein are provided as part of a pharmaceutical composition. For example, some embodiments provide pharmaceutical compositions comprising a Cas9 protein comprising an intein, or fusion proteins comprising a dCas9 protein with an intein fused to a nuclease,
recombinase, deaminase, or a transcriptional activator as provided herein, or a nucleic acid encoding such a protein, and a pharmaceutically acceptable excipient. Pharmaceutical compositions may further comprise one or more gRNA(s).

[0097] In some embodiments, compositions provided herein are administered to a subject, for example, to a human subject, in order to effect a targeted genomic modification within the subject. In some embodiments, cells are obtained from the subject and are contacted with an inventive ligand-dependent site-specific enzyme ex vivo. In some embodiments, cells removed from a subject and contacted ex vivo with an inventive ligand-dependent site-specific enzyme are re-introduced into the subject, optionally after the desired genomic modification has been effected and/or detected in the cells. Although the descriptions of pharmaceutical compositions provided herein are principally directed to pharmaceutical compositions which are suitable for administration to humans, it will be understood by the skilled artisan that such compositions are generally suitable for administration to animals of all sorts. Modification of pharmaceutical compositions suitable for administration to humans in order to render the compositions suitable for administration to various animals is well understood, and the ordinarily skilled veterinary pharmacologist can design and/or perform such modification with merely ordinary, if any, experimentation. Subjects to which administration of the pharmaceutical compositions is contemplated include, but are not limited to, humans and/or other primates; mammals, domesticated animals, pets, and commercially relevant mammals such as cattle, pigs, horses, sheep, cats, dogs, mice, and/or rats; and/or birds, including commercially relevant birds such as chickens, ducks, geese, and/or turkeys.

[0098] Formulations of the pharmaceutical compositions described herein may be prepared by any method known or hereafter developed in the art of pharmacology. In general, such preparatory methods include the step of bringing the active ingredient into association with an excipient, and then, if necessary and/or desirable, shaping and/or packaging the product into a desired single- or multi-dose unit.

[0099] Pharmaceutical formulations may additionally comprise a pharmaceutically acceptable excipient, which, as used herein, includes any and all solvents, dispersion media, diluents, or other liquid vehicles, dispersion or suspension aids, surface active agents, isotonic agents, thickening or emulsifying agents, preservatives, solid binders, lubricants and the like, as suited to the particular dosage form desired. Remington's The Science and Practice of Pharmacy, 21st Edition, A. R. Gennaro (Lippincott, Williams & Wilkins, Baltimore, MD, 2006; incorporated in its entirety herein by reference) discloses various excipients used in
formulating pharmaceutical compositions and known techniques for the preparation thereof. See also PCT application PCT/US20 10/055 131, incorporated in its entirety herein by reference, for additional suitable methods, reagents, excipients and solvents for producing pharmaceutical compositions comprising a nuclease. Except insofar as any conventional excipient medium is incompatible with a substance or its derivatives, such as by producing any undesirable biological effect or otherwise interacting in a deleterious manner with any other component(s) of the pharmaceutical composition, its use is contemplated to be within the scope of this disclosure.

[00100] In some embodiments, compositions in accordance with the present invention may be used for treatment of any of a variety of diseases, disorders, and/or conditions, including, but not limited to, autoimmune disorders (e.g. diabetes, lupus, multiple sclerosis, psoriasis, rheumatoid arthritis); inflammatory disorders (e.g. arthritis, pelvic inflammatory disease); infectious diseases (e.g. viral infections (e.g., HIV, HCV, RSV), bacterial infections, fungal infections, sepsis); neurological disorders (e.g. Alzheimer’s disease, Huntington’s disease; autism; Duchenne muscular dystrophy); cardiovascular disorders (e.g. atherosclerosis, hypercholesterolemia, thrombosis, clotting disorders, angiogenic disorders such as macular degeneration); proliferative disorders (e.g. cancer, benign neoplasms); respiratory disorders (e.g. chronic obstructive pulmonary disease); gastrointestinal disorders (e.g. inflammatory bowel disease, ulcers); musculoskeletal disorders (e.g. fibromyalgia, arthritis); endocrine, metabolic, and nutritional disorders (e.g. diabetes, osteoporosis); genitourinary disorders (e.g. renal disease); psychological disorders (e.g. depression, schizophrenia); skin disorders (e.g. wounds, eczema); and blood and lymphatic disorders (e.g. anemia, hemophilia); etc.

Methods

[00101] In another aspect of this disclosure, methods for site-specific nucleic acid (e.g., DNA) modification are provided. In some embodiments, the methods comprise contacting a DNA with any of the ligand-dependent Cas9 proteins (complexed with a gRNA) described herein, either before or after contacting the protein with a ligand that induces self-excision of the ligand-dependent intein thereby activating the nuclease. For example, in some embodiments, the method comprises (a) contacting a RNA-guided nuclease (e.g., a Cas9 protein including Cas9 nickase) comprising a ligand-dependent intein with a ligand, wherein binding of the ligand to the intein induces self-excision of the intein; and (b); contacting a DNA with the RNA-guided nuclease, wherein the RNA-guided nuclease is
associated with a gRNA; whereby self-excision of the intein from the RNA-guided nuclease in step (a) allows the RNA-guided nuclease to cleave the DNA, thereby producing cleaved DNA. In some embodiments, for examples those involving the use of an intein containing Cas9 nickase, the method produces a single strand break in the DNA. In some embodiments, the method produces a double strand break in the DNA. In some embodiments, the RNA-guided nuclease is able to bind a gRNA and optionally bind a target nucleic acid prior to being contacted with a ligand that induces self-excision of the intein, but the RNA-guided nuclease is unable to cleave the target nucleic acid until self-excision of the intein occurs. In some embodiments, the RNA-guided nuclease is unable to bind a gRNA and therefore is unable to bind a target nucleic acid until the RNA-guided nuclease is contacted with a ligand that induces self-excision of the intein. In some embodiments, the RNA-guided nuclease is any nuclease comprising Cas9 (or a variant or a fragment thereof) which comprises a ligand-dependent intein as provided herein.

[00102] In some embodiments, the method involves the use of fusion proteins comprising a nuclease-inactivated Cas9 (e.g., dCas9) fused to a nuclease domain (e.g., FokI), wherein the fusion protein comprises a ligand-dependent intein (e.g., in the dCas9 domain as provided herein), and the fusion protein lacks one or more activities (as described herein) prior to excision of the intein. In some embodiments, the fusion protein is any fusion protein described herein. In some embodiments, the method comprises contacting a target nucleic acid (e.g., DNA) with two such fusion proteins, each comprising a distinct gRNA that targets the nucleic acid, and the method comprises contacting the target nucleic acid with two such fusion proteins. The method increases the specificity of cleavage, and therefore decreases off target effects, as two fusions are required to bind the target site to elicit any nuclease activity as the nuclease domains fused to the dCas9 domain typically must dimerize at the target site to induce cleavage. In some embodiments, the method comprises contacting the fusion proteins with a ligand that induces self-excision of the intein, either before or after the gRNAs bind the fusion proteins, and/or before or after the fusion proteins bind the target nucleic acid. Once the fusion proteins are activated following excision of the intein, the nuclease domains (e.g., the FokI domains) dimerize and cleave and the target nucleic acid. Compositions and methods of using dCas9-FokI fusions are known to those of skill in the art (see, e.g., U.S. Patent Application No.: 14/320,498, titled "CAS9 VARIANTS AND USES THEREOF" which was filed on June 30, 2014; the entire contents of which are incorporated herein by reference). Those of skill in the art are routinely able to design appropriate gRNAs that target two of the fusion proteins to a target nucleic acid, and understand that in some
aspects the gRNAs are designed to hybridize to regions of the target nucleic acid that are no more than 10, no more than 15, no more than 20, no more than 25, no more than 30, no more than 40, no more than 50, no more than 60, no more than 70, no more than 80, no more than 90, or no more than 100 base pairs apart.

[00103] In some embodiments, any of the methods provided herein can be performed on DNA in a cell. For example, in some embodiments the DNA contacted by any of the inventive ligand-dependent site-specific Cas9 enzymes provided herein is in a eukaryotic cell. In some embodiments, the eukaryotic cell is an individual. In some embodiments, the individual is a human. In some embodiments, any of the methods provided herein are performed in vitro. In some embodiments, any of the methods provided herein are performed in vivo.

[00104] In some embodiments of this disclosure, methods for site-specific nucleic acid (e.g., DNA) recombination are provided. In some embodiments, the methods are useful for inducing recombination of or between two or more regions of two or more nucleic acids (e.g., DNA). In some embodiments, the methods are useful for inducing recombination of or between two or more regions in a single nucleic acid molecule (e.g., DNA). Because the recombinase fusion proteins used in the methods are ligand-dependent, the timing of recombination can be controlled to minimize off-target effects. In some embodiments, the recombination of one or more target nucleic acid molecules requires the formation of a tetrameric complex at the target site. Typically, the tetramer comprises four (4) inventive RNA-guided recombinase fusion proteins (e.g., a complex of any four inventive recombinase fusion proteins provided herein). In some embodiments, each recombinase fusion protein of the tetramer targets a particular DNA sequence via a distinct gRNA bound to each recombinase fusion protein. In some embodiments, the fusion proteins are first contacted with a ligand that induces self-excision of the intein, thereby allowing the fusion proteins to (i) bind a gRNA, (ii) bind a target nucleic acid(s), and (iii) form a complex to induce recombination between the target nucleic acid(s). In some embodiments, the fusion proteins are able to bind a gRNA prior to excision of the intein and optionally are able to bind the target nucleic acid(s) but are unable to induce recombination until the intein is excised (e.g., through the addition of a ligand that binds the ligand-dependent intein). Any of the ligand-dependent recombinase fusion proteins provided herein are useful for methods for site-specific recombination.

[00105] In some embodiments, the method for site-specific recombination between two DNA molecules comprises (a) contacting a first DNA with a first ligand-dependent
RNA-guided recombinase fusion protein, wherein the nuclease-inactivated Cas9 domain binds a first gRNA that hybridizes to a region of the first DNA; (b) contacting the first DNA with a second ligand-dependent RNA-guided recombinase fusion protein, wherein the nuclease-inactivated Cas9 domain of the second fusion protein binds a second gRNA that hybridizes to a second region of the first DNA; (c) contacting a second DNA with a third ligand-dependent RNA-guided recombinase fusion protein, wherein the nuclease-inactivated Cas9 domain of the third fusion protein binds a third gRNA that hybridizes to a region of the second DNA; and (d) contacting the second DNA with a fourth ligand-dependent RNA-guided recombinase fusion protein, wherein the nuclease-inactivated Cas9 domain of the fourth fusion protein binds a fourth gRNA that hybridizes to a second region of the second DNA. In some embodiments, the fusion proteins are first contacted with a ligand that induces self-excision of the intein prior to forming a complex with a gRNA and/or prior to hybridizing with a target DNA. In some embodiments, the method comprises contacting the fusion proteins with the ligand after the fusion proteins form a complex and/or hybridizes to a target DNA. Typically, the binding of the fusion proteins in steps (a) - (d) results in the tetramerization of the recombinase catalytic domains of the fusion proteins, such that the DNAs are recombined (i.e., following excision of the intein). In some embodiments, the gRNAs of steps (a) and (b) hybridize to opposing strands of the first DNA, and the gRNAs of steps (c) and (d) hybridize to opposing strands of the second DNA. In some embodiments, the target sites of the gRNAs of steps (a) - (d) are spaced to allow for tetramerization of the recombinase catalytic domains. For example, in some embodiments, the target sites of the gRNAs of steps (a) - (d) are no more than 10, no more 15, no more than 20, no more than 25, no more than 30, no more than 40, no more than 50, no more than 60, no more than 70, no more than 80, no more than 90, or no more than 100 base pairs apart. In some embodiments, the two regions of the two DNA molecules being recombined share homology, such that the regions being recombined are at least 80%, at least 90%, at least 95%, at least 98%, or are 100% homologous.

In some embodiments, methods for site-specific recombination between two regions of a single DNA molecule are provided. In some embodiments, the methods comprise (a) contacting a DNA with a first dCas9-recombinase fusion protein, wherein the dCas9 domain binds a first gRNA that hybridizes to a region of the DNA; (b) contacting the DNA with a second dCas9-recombinase fusion protein, wherein the dCas9 domain of the second fusion protein binds a second gRNA that hybridizes to a second region of the DNA; (c) contacting the DNA with a third dCas9-recombinase fusion protein, wherein the dCas9
domain of the third fusion protein binds a third gRNA that hybridizes to a third region of the
DNA; and (d) contacting the DNA with a fourth dCas9-recombinase fusion protein, wherein
the dCas9 domain of the fourth fusion protein binds a fourth gRNA that hybridizes to a fourth
region of the DNA. In some embodiments, the fusion proteins are first contacted with a
ligand that induces self-excision of the intein prior to forming a complex with a gRNA and/or
prior to hybridizing with a target DNA. In some embodiments, the method comprises
contacting the fusion proteins with the ligand after the fusion proteins form a complex and/or
hybridizes to a target DNA. Typically, the binding of the fusion proteins in steps (a) - (d)
results in the tetramerization of the recombinase catalytic domains of the fusion proteins,
such that the DNA is recombined (e.g. following the excision of the intein). In some
embodiments, two of the gRNAs of steps (a) - (d) hybridize to the same strand of the DNA,
and the other two gRNAs of steps (a)- (d) hybridize to the opposing strand of the DNA. In
some embodiments, the gRNAs of steps (a) and (b) hybridize to regions of the DNA that are
no more 10, no more than 15, no more than 20, no more than 25, no more than 30, no more
than 40, no more than 50, no more than 60, no more than 70, no more than 80, no more than
90, or no more than 100 base pairs apart, and the gRNAs of steps (c) and (d) hybridize to
regions of the DNA that are no more than 10, no more 15, no more than 20, no more than 25,
no more than 30, no more than 40, no more than 50, no more than 60, no more than 70, no
more than 80, no more than 90, or no more than 100 base pairs apart. In some embodiments,
the two regions of the DNA molecule being recombined share homology, such that the
regions being recombined are at least 80%, at least 90%, at least 95%, at least 98%, or are
100% homologous.

[00107] In some embodiments, any of the inventive methods for site-specific
recombination are amenable for inducing recombination, such that the recombination results
in excision (e.g., a segment of DNA is excised from a target DNA molecule), insertion (e.g.,
a segment of DNA is inserted into a target DNA molecule), inversion (e.g., a segment of
DNA is inverted in a target DNA molecule), or translocation (e.g., the exchange of DNA
segments between one or more target DNA molecule(s)). In some embodiments, the
particular recombination event (e.g., excision, insertion, inversion, translocation, etc.)
depends, inter alia, on the orientation (e.g., with respect to the target DNA molecule(s)) of
the bound RNA-guided recombinase fusion protein(s). In some embodiments, the
orientation, or direction, in which a RNA-guided recombinase fusion protein binds a target
nucleic acid can be controlled, e.g., by the particular sequence of the gRNA bound to the
RNA-guided recombinase fusion protein(s). Methods for controlling or directing a particular
recombination event are known in the art, and include, for example, those described by Turan and Bode, "Site-specific recombinases: from tag-and-target- to tag-and-exchange-based genomic modifications." *FASEB J.* 2011; Dec;25(12):4088-107, the entire contents of which are hereby incorporated by reference.

[00108] In some embodiments, any of the methods for site-specific recombination can be performed *in vivo* or *in vitro*. In some embodiments, any of the methods for site-specific recombination are performed in a cell (e.g., recombining genomic DNA in a cell). The cell can be prokaryotic or eukaryotic. The cell, such as a eukaryotic cell, can be in an individual, such as a subject, as described herein (e.g., a human subject). The methods described herein are useful for the genetic modification of cells *in vitro* and *in vivo*, for example, in the context of the generation of transgenic cells, cell lines, or animals, or in the alteration of genomic sequence, *e.g.*, the correction of a genetic defect, in a cell in or obtained from a subject. In some embodiments, a cell obtained from a subject and modified according to the methods provided herein, is re-introduced into a subject (e.g., the same subject), *e.g.*, to treat a disease, or for the production of genetically modified organisms in agricultural, medical, or biological research.

[00109] In applications in which it is desirable to recombine two or more nucleic acids so as to insert a nucleic acid sequence into a target nucleic acid, a nucleic acid comprising a donor sequence to be inserted is also provided, *e.g.*, to a cell. By a "donor sequence" it is meant a nucleic acid sequence to be inserted at the target site induced by one or more RNA-guided recombinase fusion protein(s). In some embodiments, *e.g.*, in the context of genomic modifications, the donor sequence will share homology to a genomic sequence at the target site, *e.g.*, 1%, 2%, 3%, 4%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, or 100% homology with the nucleotide sequences flanking the target site, *e.g.*, within about 100 bases or less of the target site, *e.g.* within about 90 bases, within about 80 bases, within about 70 bases, within about 60 bases, within about 50 bases, within about 40 bases, within about 30 bases, within about 15 bases, within about 10 bases, within about 5 bases, or immediately flanking the target site. In some embodiments, the donor sequence does not share any homology with the target nucleic acid, *e.g.*, does not share homology to a genomic sequence at the target site. Donor sequences can be of any length, *e.g.*, 10 nucleotides or more, 50 nucleotides or more, 100 nucleotides or more, 250 nucleotides or more, 500 nucleotides or more, 1000 nucleotides or more, 5000 nucleotides or more, 10000 nucleotides or more, 100000 nucleotides or more, *etc.*
[00110] Typically, the donor sequence is not identical to the target sequence that it replaces or is inserted into. In some embodiments, the donor sequence contains at least one or more single base changes, insertions, deletions, inversions, or rearrangements with respect to the target sequence (e.g., target genomic sequence). In some embodiments, donor sequences also comprise a vector backbone containing sequences that are not homologous to the DNA region of interest and that are not intended for insertion into the DNA region of interest.

[00111] The donor sequence may comprise certain sequence differences as compared to the target (e.g., genomic) sequence, for example, restriction sites, nucleotide polymorphisms, selectable markers (e.g., drug resistance genes, fluorescent proteins, enzymes etc.), which can be used to assess successful insertion of the donor sequence at the target site or in some cases may be used for other purposes (e.g., to signify expression at the targeted genomic locus). In some embodiments, if located in a coding region, such nucleotide sequence differences will not change the amino acid sequence, or will make silent amino acid changes (e.g., changes which do not affect the structure or function of the protein). In some embodiments, these sequences differences may include flanking recombination sequences such as FLPs, loxP sequences, or the like, that can be activated at a later time for removal of e.g., a marker sequence. The donor sequence may be provided to the cell as single-stranded DNA, single-stranded RNA, double-stranded DNA, or double-stranded RNA. It may be introduced into a cell in linear or circular form. If introduced in linear form, the ends of the donor sequence may be protected (e.g., from exonuclease degradation) by methods known to those of skill in the art. For example, one or more dideoxynucleotide residues are added to the 3' terminus of a linear molecule and/or self-complementary oligonucleotides are ligated to one or both ends. See, e.g., Chang et al., Proc. Natl. Acad Sci USA. 1987; 84:4959-4963; Nehls et al., Science. 1996; 272:886-889. In some embodiments, a donor sequence is introduced into a cell as part of a vector molecule having additional sequences such as, for example, replication origins, promoters, and genes encoding antibiotic resistance. In some embodiments, donor sequences can be introduced as naked nucleic acid, as nucleic acid complexed with an agent such as a liposome or polymer (e.g., poloxamer), or can be delivered by viruses (e.g., adenovirus, AAV, etc.).

[00112] In some embodiments, any of the methods provided herein can be performed on DNA in a cell. For example, in some embodiments the DNA contacted by any RNA/gPvNA-comprising complex provided herein is in a eukaryotic cell. In some embodiments, the eukaryotic cell is in an individual. In some embodiments, the individual is
a human. In some embodiments, any of the methods provided herein are performed in vitro. In some embodiments, any of the methods provided herein are performed in vivo.

[00113] In some embodiments of this disclosure, methods for site-specific nucleic acid (e.g., DNA) editing are provided. In some embodiments, the fusion protein is used to introduce a point mutation into a nucleic acid by deaminating a target nucleobase, e.g., a cytidine (C) residue. In some embodiments, the method comprises contacting a DNA molecule with a ligand-dependent fusion protein comprising a nuclease inactivated RNA-guided nuclease (e.g., dCas9), which comprises a ligand dependent intein, fused to a deaminase, and (b) a gRNA targeting the fusion protein of step (a) to a target nucleotide sequence of the DNA strand; wherein the DNA molecule is contacted with the fusion protein and the gRNA in an amount effective and under conditions suitable for the deamination of a nucleotide base. Any of the fusion proteins comprising a gene editing domain as provided herein are amenable for use in the methods. In some embodiments, the method first comprises contacting the fusion protein with a ligand that induces self-excision of the intein prior to forming a complex with the gRNA. In some embodiments, the method comprises contacting the fusion protein with a ligand that induces self-excision of the intein after the fusion protein has formed a complex with the gRNA.

[00114] In some embodiments, the deamination of the target nucleobase results in the correction of a genetic defect, e.g., in the correction of a point mutation that leads to a loss of function in a gene product. In some embodiments, the genetic defect is associated with a disease or disorder, e.g., a lysosomal storage disorder or a metabolic disease, such as, for example, type I diabetes. In some embodiments, the methods provided herein are used to introduce a deactivating point mutation into a gene or allele that encodes a gene product that is associated with a disease or disorder. For example, in some embodiments, methods are provided herein that employ a DNA editing fusion protein to introduce a deactivating point mutation into an oncogene (e.g., in the treatment of a proliferative disease). A deactivating mutation may, in some embodiments, generate a premature stop codon in a coding sequence, which results in the expression of a truncated gene product, e.g., a truncated protein lacking the function of the full-length protein.

[00115] In some embodiments, the purpose of the methods provided herein is to restore the function of a dysfunctional gene via genome editing. Compositions and methods of using gene editing enzymes fused e.g., to Cas9 are known, and include those described in U.S. Patent Application No.: 14/325,815 titled "FUSIONS OF CAS9 DOMAINS AND NUCLEIC ACID-EDITING DOMAINS," and filed on July 8, 2014; the entire contents of
which are incorporated herein by reference. The fusion proteins provided herein (comprising ligand-dependent inteins) can be validated for gene editing-based human therapeutics in vitro, e.g., by correcting a disease-associated mutation in human cell culture. It will be understood by the skilled artisan that the fusion proteins provided herein, e.g., the fusion proteins comprising a dCas9 domain (e.g., comprising a ligand-dependent intein) and a nucleic acid deaminase domain can be used to correct any single point T -> C or A -> G mutation. In the first case, deamination of the mutant C back to U corrects the mutation, and in the latter case, deamination of the C that is base-paired with the mutant G, followed by a round of replication, corrects the mutation.

[00116] An exemplary disease-relevant mutation that can be corrected by the provided fusion proteins in vitro or in vivo is the H1047R (A3140G) polymorphism in the PI3KCA protein. The phosphoinositide-3-kinase, catalytic alpha subunit (PI3KCA) protein acts to phosphorylate the 3-OH group of the inositol ring of phosphatidylinositol. The PI3KCA gene has been found to be mutated in many different carcinomas, and thus it is considered to be a potent oncogene (Lee et al., PIK3CA gene is frequently mutated in breast carcinomas and hepatocellular carcinomas. Oncogene. 2005; 24(8): 1477-80). In fact, the A3140G mutation is present in several NCT60 cancer cell lines, such as, for example, the HCT16, SKOV3, and T47D cell lines, which are readily available from the American Type Culture Collection (ATCC)(Ikediobi et al., Mutation analysis of 24 known cancer genes in the NCT60 cell line set. Mol Cancer Ther. 2006; 5(11):2606-12.

[00117] In some embodiments, a cell carrying a mutation to be corrected, e.g., a cell carrying a point mutation, e.g., an A3140G point mutation in exon 20 of the PI3KCA gene, resulting in a H1047R substitution in the PI3KCA protein is contacted with an expression construct encoding a ligand-dependent Cas9 deaminase fusion protein and an appropriately designed gRNA targeting the fusion protein to the respective mutation site in the encoding PI3KCA gene. Control experiments can be performed where the gRNAs are designed to target the fusion enzymes to non-C residues that are within the PI3KCA gene. Genomic DNA of the treated cells can be extracted, and the relevant sequence of the PI3KCA genes PCR amplified and sequenced to assess the activities of the fusion proteins in human cell culture.

[00118] It will be understood that the example of correcting point mutations in PI3KCA is provided for illustration purposes and is not meant to limit the instant disclosure. The skilled artisan will understand that the instantly disclosed ligand-dependent DNA-editing fusion proteins can be used to correct other point mutations and mutations associated with other cancers and with diseases other than cancer including other proliferative diseases.
The successful correction of point mutations in disease-associated genes and alleles opens up new strategies for gene correction with applications in therapeutics and basic research. Site-specific single-base modification systems like the disclosed fusions of ligand-dependent Cas9 and deaminase enzymes or domains also have applications in "reverse" gene therapy, where certain gene functions are purposely suppressed or abolished. In these cases, site-specifically mutating Trp (TGG), Gin (CAA and CAG), or Arg (CGA) residues to premature stop codons (TAA, TAG, TGA) can be used to abolish protein function in vitro, ex vivo, or in vivo.

The instant disclosure provides methods for the treatment of a subject diagnosed with a disease associated with or caused by a point mutation that can be corrected by a ligand-dependent Cas9 DNA editing fusion protein provided herein. For example, in some embodiments, a method is provided that comprises administering to a subject having such a disease, e.g., a cancer associated with a PI3KCA point mutation as described above, an effective amount of a ligand-dependent Cas9 deaminase fusion protein that corrects the point mutation or introduces a deactivating mutation into the disease-associated gene, e.g., following subsequent administration of the small molecule (e.g., ligand) that activates the fusion protein. In some embodiments, the disease is a proliferative disease. In some embodiments, the disease is a genetic disease. In some embodiments, the disease is a neoplastic disease. In some embodiments, the disease is a metabolic disease. In some embodiments, the disease is a lysosomal storage disease. Other diseases that can be treated by correcting a point mutation or introducing a deactivating mutation into a disease-associated gene will be known to those of skill in the art, and the disclosure is not limited in this respect.

The instant disclosure provides methods for the treatment of additional diseases or disorders, e.g., diseases or disorders that are associated or caused by a point mutation that can be corrected by deaminase-mediated gene editing. Some such diseases are described herein, and additional suitable diseases that can be treated with the strategies and fusion proteins provided herein will be apparent to those of skill in the art based on the instant disclosure. Exemplary suitable diseases and disorders include, without limitation, cystic fibrosis (see, e.g., Schwank et al., Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. Cell stem cell. 2013; 13: 653-658; and Wu et al., Correction of a genetic disease in mouse via use of CRISPR-Cas9. Cell stem cell. 2013; 13: 659-662, neither of which uses a deaminase fusion protein to correct the genetic defect); phenylketonuria - e.g., phenylalanine to serine mutation at position 835 in...
phenylalanine hydroxylase gene (T>C mutation) - see, e.g., McDonald et al., Genomics. 1997; 39:402-405; Bernard-Soulier syndrome (BSS) - e.g., phenylalanine to serine mutation at position 55 in the platelet membrane glycoprotein IX (T>C mutation) - see, e.g., Noris et al., British Journal of Haematology. 1997; 97: 312-320; epidermolytic hyperkeratosis (EHK) - e.g., leucine to proline mutation at position 160 in keratin 1 (T>C mutation) - see, e.g., Chipev et al., Cell. 1992; 70: 821-828; chronic obstructive pulmonary disease (COPD) - e.g., leucine to proline mutation at position 55 in α1-antitrypsin (T>C mutation) - see, e.g., Poller et al., Genomics. 1993; 17: 740-743; Charcot-Marie-Tooth disease type 4J - e.g., leucine to proline mutation at position 197 in FIG4 (T>C mutation) - see, e.g., Kundu et al., J Biotechnol. 2013; 3: 225-234; neuroblastoma (NB) - e.g., isoleucine to threonine mutation at position 41 in Caspase-9 (T>C mutation) - see, e.g., Lenk et al., PLoS Genetics. 2011; 7: e1002104; von Willebrand disease (vWD) - e.g., cysteine to arginine mutation at position 509 in von Willebrand factor (T>C mutation) - see, e.g., Lavergne et al., Br. J. Haematol. 1992; 82: 66-72; myotonia congenital - e.g., cysteine to arginine mutation at position 277 in the muscle chloride channel gene CLCN1 (T>C mutation) - see, e.g., Weinberger et al., The J. of Physiology. 2012; 590: 3449-3464; hereditary renal amyloidosis - e.g., stop codon to arginine mutation at position 78 in apolipoprotein All (T>C mutation) - see, e.g., Yazaki et al., Kidney Int. 2003; 64: 11-16; dilated cardiomyopathy (DCM) - e.g., tryptophan to Arginine mutation at position 148 in the FOXD4 gene (T>C mutation), see, e.g., Minoretti et. al., Int. J. of Mol. Med. 2007; 19: 369-372; hereditary lymphedema - e.g., histidine to arginine mutation at position 1035 in tyrosine kinase (A>G mutation), see, e.g., Irthum et al., Am. J. Hum. Genet. 2000; 67: 295-301; familial Alzheimer's disease - e.g., isoleucine to valine mutation at position 143 in presenilin1 (A>G mutation), see, e.g., Gallo et. al., J. Alzheimer's disease. 2011; 25: 425-431; Prion disease - e.g., methionine to valine mutation at position 129 in prion protein (A>G mutation) - see, e.g., Lewis et. al., J. of General Virology. 2006; 87: 2443-2449; chronic infantile neurologic cutaneous articular syndrome (CINCA) - e.g., Tyrosine to Cysteine mutation at position 570 in cryopyrin (A>G mutation) - see, e.g., Fujisawa et. al. Blood. 2007; 109: 2903-2911; and desmin-related myopathy (DRM) - e.g., arginine to glycine mutation at position 120 in B crystallin (A>G mutation) - see, e.g., Kumar et al., J. Biol. Chem. 1999; 274: 24137-24141. The entire contents of each of the foregoing references and database entries are incorporated herein by reference.

According to another aspect, methods for transcriptional activation of a gene are provided. In some embodiments, the methods comprise contacting a DNA molecule
comprising a gene with (a) a ligand-dependent dCas9 fusion protein comprising a transcriptional activator (e.g., any of those provided herein) and (b) a gRNA targeting the fusion protein of (a) to a target nucleotide sequence of the DNA strand; wherein the DNA molecule is contacted with the fusion protein and the gRNA in an amount effective and under conditions suitable for the transcriptional activation of the gene. In some embodiments, the method further comprises contacting the fusion protein with a ligand that induces self-excision of the intein. In some embodiments, the fusion protein is contacted with the ligand prior to forming a complex with a gRNA. In some embodiments, the fusion protein is contacted with the ligand after forming a complex with a gRNA. In some embodiments, the gRNA targets the promoter region of a gene. Methods for inducing gene activation using fusion proteins comprising a transcriptional activator are known in the art, and include those described by Perex-Pinera et al., "RNA-guided gene activation by CRISPR-Cas9-based transcription factors." Nature Methods. 2013; 10, 973-976; the entire contents of which are incorporated herein by reference.

According to another aspect, methods for transcriptional repression of a gene are provided. In some embodiments, the methods comprise contacting a DNA molecule comprising a gene with (a) a ligand-dependent dCas9 fusion protein comprising a transcriptional repressor (e.g., any of those provided herein) and (b) a gRNA targeting the fusion protein of (a) to a target nucleotide sequence of the DNA strand; wherein the DNA molecule is contacted with the fusion protein and the gRNA in an amount effective and under conditions suitable for the transcriptional repression of the gene. In some embodiments, the method further comprises contacting the fusion protein with a ligand that induces self-excision of the intein. In some embodiments, the fusion protein is contacted with the ligand prior to forming a complex with a gRNA. In some embodiments, the fusion protein is contacted with the ligand after forming a complex with a gRNA. In some embodiments, the gRNA targets the promoter region of a gene. Methods for inducing gene repression using fusion proteins comprising a transcriptional repressor are known in the art, and include those described by Gilbert et al., CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell. 2013; 154, 442-451; the entire contents of which are incorporated herein by reference.

According to another aspect, methods for epigenetic modification of DNA are provided. In some embodiments, the methods comprise contacting a DNA molecule comprising with (a) a ligand-dependent dCas9 fusion protein comprising an epigenetic modifier (e.g., any of those provided herein) and (b) a gRNA targeting the fusion protein of
to a target nucleotide sequence of the DNA strand; wherein the DNA molecule is contacted with the fusion protein and the gRNA in an amount effective and under conditions suitable for the epigenetic modification of the DNA. In some embodiments, the DNA comprises one or more genes. In some embodiments, the method further comprises contacting the fusion protein with a ligand that induces self-excision of the intein. In some embodiments, the fusion protein is contacted with the ligand prior to forming a complex with a gRNA. In some embodiments, the fusion protein is contacted with the ligand after forming a complex with a gRNA. In some embodiments, the gRNA targets the promoter region of a gene. In some embodiments, the epigenetic modification that results is methylation of DNA. In some embodiments, the epigenetic modification that results is demethylation of DNA. In some embodiments, the epigenetic modification that results is methylation of histone protein. In some embodiments, the epigenetic modification that results is demethylation of histone protein. In some embodiments, the epigenetic modification that results is acetylation of histone protein. In some embodiments, the epigenetic modification that results is deacetylation of histone protein. Methods for inducing epigenetic modifications using fusion proteins comprising an epigenetic modifier are known in the art, and include those described by Konermann et al., Optical control of mammalian endogenous transcription and epigenetic states. Nature. 2013; 500, 472-476; Mendenhall et al., Locus-specific editing of histone modifications at endogenous enhancers. Nat. Biotechnol. 2013; 31, 1133-1136; and Maeder et al., Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. Nat. Biotechnol. 2013; 31, 1137-1142; the entire contents of which are incorporated herein by reference.

In some embodiments, any of the methods provided herein can be performed on DNA in a cell. For example, in some embodiments the DNA contacted by any RNA/gRNA-comprising complex provided herein is in a eukaryotic cell. In some embodiments, the eukaryotic cell is in an individual. In some embodiments, the individual is a human. In some embodiments, any of the methods provided herein are performed in vitro. In some embodiments, any of the methods provided herein are performed in vivo.

In some embodiments of the methods provided herein, the ligand-dependent Cas9 protein, e.g., the Cas9-intein or the Cas9-intein fusion protein, is contacted with the ligand at a concentration effective to excise the intein from the Cas9-intein variant, or at a concentration effective to induce a desired modification (e.g., cleavage, nicking, recombination, or deamination) of a target site. In some embodiments, a ligand-dependent
Cas9 protein provided herein is contacted with a suitable ligand at a concentration resulting in decreased off-target activity of the Cas9 protein as compared to the off-target activity of wild-type Cas9. For example, in some embodiments, a method provided herein comprises contacting a population of ligand-dependent Cas9 proteins in vitro or in vivo in the presence of a target nucleic acid to be modified with a suitable ligand at a concentration resulting in the desired modification of the target nucleic acid, and in either no off-target activity (i.e., no modification of any non-target nucleic acids) or in an off-target activity of less than 80%, less than 75%, less than 50%, less than 45%, less than 40%, less than 35%, less than 30%, less than 25%, less than 24%, less than 23%, less than 22%, less than 21%, less than 20, less than 19%, less than 18%, less than 17%, less than 16%, less than 15%, less than 14%, less than 13%, less than 12%, less than 11%, less than 10%, less than 9%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, or less than 1% of the off-target activity observed or expected under the same conditions when using wild-type Cas9.

**Polynucleotides, Vectors, Cells, Kits**

[00127] In another aspect of this disclosure, polynucleotides encoding one or more of the inventive proteins and/or gRNAs are provided. For example, polynucleotides encoding any of the proteins described herein are provided, e.g., for recombinant expression and purification of isolated nucleases, recombinases, gene editing enzymes, and other nucleic acid modifying enzymes, e.g., comprising Cas9 variants (e.g., dCas9) comprising ligand-dependent inteins. In some embodiments, an isolated polynucleotide comprises one or more sequences encoding a ligand dependent RNA-guided nuclease (e.g., Cas9). In some embodiments, an isolated polynucleotide comprises one or more sequences encoding a Cas9 fusion protein, for example, any of the Cas9 fusion proteins described herein (e.g., those comprising a nuclease-inactivated Cas9 fused to a nuclease, recombinase, deaminase domain, or transcriptional activator). In some embodiments, an isolated polynucleotides comprises one or more sequences encoding a gRNA, alone or in combination with a sequence encoding any of the proteins described herein.

[00128] In some embodiments, vectors encoding any of the proteins described herein are provided, e.g., for recombinant expression and purification of Cas9 proteins, and/or fusions comprising Cas9 proteins (e.g., variants). In some embodiments, the vector comprises or is engineered to include an isolated polynucleotide, e.g., those described herein. In some embodiments, the vector comprises one or more sequences encoding a Cas9 protein
(as described herein), a gRNA, or combinations thereof, as described herein. Typically, the vector comprises a sequence encoding an inventive protein operably linked to a promoter, such that the fusion protein is expressed in a host cell.

[00129] In some embodiments, cells are provided, e.g., for recombinant expression and purification of any of the Cas9 proteins provided herein. The cells include any cell suitable for recombinant protein expression, for example, cells comprising a genetic construct expressing or capable of expressing an inventive protein (e.g., cells that have been transformed with one or more vectors described herein, or cells having genomic modifications, for example, those that express a protein provided herein from an allele that has been incorporated into the cell's genome). Methods for transforming cells, genetically modifying cells, and expressing genes and proteins in such cells are well known in the art, and include those provided by, for example, Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)) and Friedman and Rossi, *Gene Transfer: Delivery and Expression ofDNA and RNA, A Laboratory Manual* (1st ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2006)).

[00130] Some aspects of this disclosure provide kits comprising a ligand-dependent Cas9 variant (e.g., a ligand dependent Cas9 nuclease (or nickase), and/or a ligand-dependent dCas9 variant fused to a nuclease, recombinase, deaminase, or a transcriptional activator as provided herein. In some embodiments, the kit comprises a polynucleotide encoding an inventive Cas9 variant, nuclease, recombinase, and/or deaminase e.g., as provided herein. In some embodiments, the kit comprises a vector for recombinant protein expression, wherein the vector comprises a polynucleotide encoding any of the proteins provided herein. In some embodiments, the kit comprises a cell (e.g., any cell suitable for expressing Cas9 proteins or fusions comprising Cas9 proteins, such as bacterial, yeast, or mammalian cells) that comprises a genetic construct for expressing any of the proteins provided herein. In some embodiments, any of the kits provided herein further comprise one or more gRNAs and/or vectors for expressing one or more gRNAs. In some embodiments, the kit comprises an excipient and instructions for contacting the Cas9 proteins or dCas9 fusions with the excipient to generate a composition suitable for contacting a nucleic acid with the inventive protein. In some embodiments, the composition is suitable for delivering an inventive protein to a cell, or for delivering a nucleic acid encoding the protein to a cell. In some embodiments, the composition is suitable for delivering an inventive protein to a subject. In some embodiments, the excipient is a pharmaceutically acceptable excipient.
EXAMPLES

Example 1: Small molecule-controlled Cas9

Cas9 variants that can be activated in the presence of a small molecule were engineered, allowing spatiotemporal control over DNA cleavage. These engineered Cas9 variants contain a small-molecule-regulated intein (Buskirk et al., Proc. Natl. Acad. Sci. USA. 2004; 101, 10505-10510), which has been optimized for mammalian cells (Peck et al., Chem. Biol. 2011; 18 (5), 619-630), that renders the protein inactive as a nuclease. Upon addition of the cell-permeable molecule, 4-hydroxytamoxifen (4-HT), the intein excises itself from the protein and ligates the flanking extein sequences, restoring Cas9 activity. Because these Cas9 variants can be active over a smaller time window than wild-type Cas9, the likelihood of having off-target cleavage is reduced.

The 37R3-2 intein was inserted at 15 different positions into human codon-optimized Streptococcus pyogenes Cas9 (e.g., SEQ ID NO:2). The intein was inserted in place of a single cysteine, alanine, serine, or threonine residue. Upon excision, the intein leaves a cysteine residue. Thus, the primary structure generated following protein splicing is either identical to the unmodified version of Cas9 when the intein is inserted in place of cysteine, or it is one amino acid different when the intein is inserted in place of alanine, serine, or threonine.

Plasmid constructs were generated in which the intein replaced amino acid residues: Cys80, Alal27, Thr46, Ser219, Thr333, Thr519, Cys574, Thr622, Ser701, Ala728, Thr995, SerL006, SerL54, SerL59, or SerL274 (e.g., in the amino acid sequence set forth as SEQ ID NO:2). These plasmids express the Cas9 variant with a nuclear localization signal (NLS) and 3xFLAG tag from the CMV promoter.

HEK293-GFP stable cells were transfected with the Cas9 expression plasmid, a gRNA (targeting Emerald GFP; Guilinger et al., Nature Biotechnology (2014)), and iRFP670 (transfection control), using Lipofectamine 2000. Twelve hours after transfection, media, either containing 4-HT (1 µM) or without 4-HT, was added.

Five days after transfection, cells were trypsinized and analyzed on a flow cytometer. Cells lacking GFP indicated genome modification. Cas9 variants that induced minimal genome modification in the absence of 4-HT but induce significant genome modification in the presence of 4-HT were deemed small-molecule-regulated variants in this Example. Of fifteen targeted insertions, five demonstrated minimal genome modification in the absence of 4-HT. These variants are highlighted in bold in the Table I below.
Additionally, a time course was performed in which incubation with 4-HT was limited to 2, 4, 8, 12 or 24 hours, after which point the media was replaced. Presumably, the shorter time an "active" cas9 is present, the less off-target cleavage. As depicted in Table II below, treating with 4-HT for 2 hours is sufficient for on-target cleavage and longer treatment periods do not show significant increased cleavage in this assay.

To assess the ability of the ligand-dependent Cas9 proteins to affect genomic modifications in the presence of absence of ligand, HEK293-GFP stable cells (GenTarget) were transfected with Cas9 expression plasmids and sgRNAs targeting EMX, VEGF, or CLTA genomic sites using Lipofectamine 2000 as previously described (Guilinger et al., Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification Nature Biotechnology. 2014; 32(6):577-82). 4-HT (1 μM) was added during transfection for + 4-HT samples. 12 hours after transfection, cells were trypsinized and genomic DNA was isolated using the DNAdvance kit (Agencourt). 40-80 ng of genomic DNA was used as a template to PCR amplify the targeted genomic loci with flanking Survey primer pairs as previously described (Guilinger et al., Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification Nature Biotechnology. 2014; 32(6):577-82). PCR products were purified with a QIAquick PCR Purification Kit (Qiagen) and quantified with a Quant-iT PicoGreen dsDNA Kit (Life Technologies). 200 ng of purified PCR DNA was then combined with 2 μL of NEBuffer 2 (NEB) in a total volume of 19 μL and denatured then re-annealed with thermocycling at 95°C for 5 min, 95-85°C at 2°C/s, 85-20°C at 0.2°C/s. The re-annealed DNA was incubated with 1 μL of T7 Endonuclease I (10U/ μL, NEB) at 37°C for 15 min. 10 μL of 50% glycerol was added to the T7 Endonuclease reaction and 15 μL was analyzed on a 5% TBE 18-well Criterion PAGE gel (Bio-Rad) electrophoresed for 30 min at 200V and stained with EtBr for 15 min.

As shown in Figure 2, the addition of 4-HT to ligand-dependent Cas9:Intein variants (Cas9:Intein with 37R3-2 replacing S219 (SEQ ID NO:30) and Cas9:Intein with 37R3-2 replacing C574 (SEQ ID NO:33)) resulted in genomic modification of the target sites, comparable to modification by wild-type Cas9. In the absence of 4-HT, the Cas9:Intein variants displayed minimal or no modification of the EMX and VEGF genomic target sites, while some background cleavage was observed for the CLTA genomic target site. Gene modification levels can be estimated by comparing the intensities of the cleaved (two smaller fragments) and uncleaved bands. These results demonstrate that Cas9 cleavage of genomic
target sites can be controlled by the addition of ligand (here, 4-HT) which activates the proteins.

Table 1

<table>
<thead>
<tr>
<th>Cas9 Variant</th>
<th>Cells without GFP (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- 4-HT</td>
</tr>
<tr>
<td>None</td>
<td>4.65</td>
</tr>
<tr>
<td>wild-type cas9</td>
<td>48.49</td>
</tr>
<tr>
<td>intein(Cys80)-Cas9 (SEQ ID NO:27)</td>
<td>7.08</td>
</tr>
<tr>
<td>intein(Ala127)-Cas9 (SEQ ID NO:28)</td>
<td>7.97</td>
</tr>
<tr>
<td>intein(Thr146)-Cas9 (SEQ ID NO:29)</td>
<td>8.77</td>
</tr>
<tr>
<td>intein(Ser219)-Cas9 (SEQ ID NO:30)</td>
<td>6.53</td>
</tr>
<tr>
<td>intein(Thr333)-Cas9 (SEQ ID NO:31)</td>
<td>4.96</td>
</tr>
<tr>
<td>intein(Thr519)-Cas9 (SEQ ID NO:32)</td>
<td>9.49</td>
</tr>
<tr>
<td>intein(Cys574)-Cas9 (SEQ ID NO:33)</td>
<td>5.74</td>
</tr>
<tr>
<td>intein(Thr622)-Cas9 (SEQ ID NO:34)</td>
<td>5.67</td>
</tr>
<tr>
<td>intein(Ser701)-Cas9 (SEQ ID NO:35)</td>
<td>6.54</td>
</tr>
<tr>
<td>intein(Ala728)-Cas9 (SEQ ID NO:36)</td>
<td>20.82</td>
</tr>
<tr>
<td>intein(Thr995)-Cas9 (SEQ ID NO:37)</td>
<td>14.95</td>
</tr>
<tr>
<td>intein(Ser1006)-Cas9 (SEQ ID NO:38)</td>
<td>6.80</td>
</tr>
<tr>
<td>intein(Ser1154)-Cas9 (SEQ ID NO:39)</td>
<td>21.14</td>
</tr>
<tr>
<td>intein(Ser1159)-Cas9 (SEQ ID NO:40)</td>
<td>5.65</td>
</tr>
<tr>
<td>intein(Ser1274)-Cas9 (SEQ ID NO:41)</td>
<td>3.08</td>
</tr>
</tbody>
</table>
Example 2: Small Molecule-controlled Cas9 Protein with Improved Genome-Editing Specificity

Cas9 nucleases that are activated by the presence of a cell-permeable small molecule were developed by inserting an evolved 4-hydroxytamoxifen (4-HT)-responsive intein at specific positions in Cas9. In human cells, conditionally active Cas9s modify target genomic sites with up to 25-fold higher specificity than wild-type Cas9.

The RNA-guided endonuclease Cas9 from the type II CRISPR-Cas system enables simple and efficient genome editing in a wide variety of organisms. Virtually any target DNA locus can be cleaved by programming Cas9 with a single guide RNA (sgRNA) that contains a stretch of ~20 nucleotides complementary to the target sequence. Due to its simplicity and robustness, the Cas9 system has been widely adopted for biological research and therapeutic development. The DNA cleavage specificity of Cas9 is imperfect, however, raising concerns over off-target genome modification that may limit its usefulness in therapeutic or research applications. Cas9 off-target activity has been reduced through
protein and sgRNA engineering, and by direct delivery of Cas9:sgRNA protein:RNA complexes into cells. 4HT was added to treat or control cells during transfection. Five days post-transfection, cells were analyzed on a

[00141] A complementary, underexplored strategy to improve Cas9 specificity is to reduce its activity once it has had sufficient opportunity to modify the target DNA locus. Indeed, higher concentrations of Cas9 in cells have been observed to degrade specificity (defined as the ratio of on-target:off-target DNA cleavage activity), presumably because any Cas9 protein present after the target locus has been modified can only process off-target substrates. Unfortunately, wild-type Cas9 nucleases are not known to be regulated by other molecules and therefore are used in constitutively active form. While Cas9 can be regulated at the transcriptional level through the use of inducible promoters, transcriptional control cannot limit activity to the short temporal windows that may be necessary to maximize genome-editing specificity, in contrast with the high temporal resolution of post-translational strategies that directly control protein activity.

[00142] Engineered variants of Cas9 that can be controlled with a readily available, cell-permeable small molecule were developed. We previously evolved inteins that undergo protein splicing only in the presence of 4-hydroxytamoxifen (4-HT). These inteins were developed by inserting the human estrogen receptor ligand-binding domain into the M. tuberculosis RecA intein and evolving the resulting inactive fusion protein into a conditionally active intein that requires the presence of 4-HT. Subsequent evolution at 37 °C yielded a second-generation intein, 37R3-2, with improved splicing properties in mammalian cells. We envisioned that inserting the 37R3-2 intein into Cas9 at a location that disrupts Cas9 activity until protein splicing has taken place could result in conditionally active Cas9 nucleases that are active only in the presence of 4-HT.

[00143] We genetically inserted the 4-HT-dependent intein at each of fifteen positions in Cas9 (Cys80, Ala27, Thr46, Ser219, Thr333, Thr519, Cys574, Thr622, Ser701, Ala728, Thr995, Ser1006, Ser154, Ser159, and Ser174), chosen to distribute the location of the intein across the structural domains of Cas9. (Fig. 3a). Because intein splicing leaves behind a single Cys residue, the intein was inserted in place of one Cas9 amino acid in each of the 15 candidate constructs. In addition to replacing natural Cys amino acids, we also favored replacing Ala, Ser, or Thr residues to minimize the likelihood that the resulting Cys point mutation resulting from protein splicing would disrupt Cas9 activity. The 15 intein-Cas9 candidates were expressed in HEK293-GFP cells together with a sgRNA that targets the genomic EGFP locus in these cells. Twelve hours post-transfection, cells were treated with or without 1 µM 4-HT. Five days post-transfection, cells were analyzed on a
flow cytometer for loss of GFP expression from Cas9-mediated EGFP cleavage and subsequent non-homologous end joining.

Eight of the candidates, corresponding to intein insertion at A127, T146, S219, T333, T519, C574, S1006, and S1159, demonstrated 4-HT-dependent loss of GFP expression consistent with 4-HT-triggered Cas9 activity (Fig. 3c). Interestingly, three intein-Cas9 proteins (insertion at A728, T995, and S1154) showed high DNA modification rates both in the presence and absence of 4-HT, suggesting that large protein insertions at these positions do not significantly inhibit nuclease activity, or that the intein lost its 4-HT dependence due to context-dependent conformational perturbations. We speculate that it may be possible to engineer split Cas9 variants by dividing the protein at these locations, given their tolerance of a 413-residue insertion. The lack of nuclease activity of the remaining four Cas9-inteins (insertion at C80, T622, S701, and S1274) in the presence or absence of 4-HT could result from the inability of the intein to splice in those contexts, the inability of Cas9 to refold properly following splicing, or intolerance of replacement of native Thr or Ser residues with Cys. We pursued two intein-Cas9 variants corresponding to insertion at S219 and C574 (Fig. 3b). These two variants combined high activity in the presence of 4-HT and low activity in the absence of 4-HT.

To evaluate the genome modification specificity of conditionally active Cas9 variants, we expressed intein-Cas9(S219), intein-Cas9(C574), and wild-type Cas9 in HEK293-GFP cells together with each of three previously described sgRNAs that target the well-studied EMX, VEGF, and CLTA genomic loci. We assayed these Cas9:sgRNA combinations in human cells for their ability to modify the three on-target loci as well as 11 known off-target genomic sites (Table 3)\textsuperscript{45,46}. Cells were treated with or without 1 µM 4-HT during transfection, and after 12 h the media was replaced with fresh media lacking 4-HT. We observed no cellular toxicity arising from 12 or 60 h of treatment with 1 µM 4-HT in untransfected or transfected HEK293 cells (Fig. 5). Genomic DNA was isolated 60 h post-transfection and analyzed by high-throughput DNA sequencing.

Overall on-target genome modification frequency of intein-Cas9(S219) and intein-Cas9 (C574) in the presence of 1 µM 4-HT was similar to that of wild-type Cas9 (Fig. 4a, Tables 4 and 5). On-target modification frequency in the presence of 4-HT was 3.4- to 7.3-fold higher for intein-Cas9(S219), and 3.6- to 9.6-fold higher for intein-Cas9(C574), than in the absence of 4-HT, whereas modification efficiency for wild-type Cas9 was 1.2- to 1.8-fold lower in the presence of 4-HT (Fig. 4a). Both intein-Cas9 variants exhibited a low level of background activity in the absence of 4-HT, consistent with previous reports\textsuperscript{20-22}. Western
blot analysis of intein-Cas9(S219) from transfected HEK293 cells confirmed the presence of spliced product at the earliest assayed time point (4 h) following 4-HT treatment; no spliced product was detected in the absence of 4-HT (Fig. 6). Together, these results indicate that intein-Cas9(S219) and intein-Cas9(C574) are slightly less active than wild-type Cas9 in the presence of 4-HT, likely due to incomplete splicing (Fig. 6), but much less active in the absence of 4-HT.

High-throughput sequencing of 11 previously described off-target sites that are modified by wild-type Cas9:sgRNA complexes targeting the EMX, VEGF, and CLTA loci revealed that both intein-Cas9 variants when treated with 4-HT for 12 h exhibit substantially improved specificity compared to that of wild-type Cas9 (Fig. 7, Tables 4, 6, and 7). On-target:off-target indel modification ratios for both intein-Cas9 variants were on average 6-fold higher, and as much as 25-fold higher, than that of wild-type Cas9 (Fig. 4b-d). In the absence of 4-HT, the genome modification specificity of both intein-Cas9 variants was on average 14-fold higher than that of wild-type Cas9 in the absence of 4-HT (Fig. 8), presumably resulting from the much lower activity of the intein-Cas9 variants in the absence of 4-HT.

Since intein-Cas9s can result in slightly lower on-target modification rates compared to wild-type Cas9 (Fig. 4a), we sought to verify that the improvements in specificity among the intein-Cas9s were not simply a result of reduced activity. Both on- and off-target activity of Cas9 has been shown to be dependent on the amount of Cas9 expression plasmid transfected. By transfecting lower amounts of the wild-type Cas9 expression plasmid, we compared intein-Cas9s with wild-type Cas9 under conditions that result in very similar levels of on-target modification. To minimize potential differences in transfection efficiency, we supplemented with a plasmid that does not express Cas9 so that the same total amount of plasmid DNA was transfected into each sample. High-throughput sequencing revealed that wild-type Cas9 shows slightly improved specificity, as expected, as the on-target cleavage rate is reduced. The intein-Cas9 variants, however, remain substantially more specific than wild-type Cas9 at similar on-target DNA cleavage rates (Figs. 9-11, Tables 6 and 8). For example, intein-Cas9(C574) and wild-type Cas9 (80 ng) have virtually identical on-target DNA cleavage rates (both 6.4%) at the EMX locus but all four off-target sites are modified at an average of 4-fold lower frequencies ($P < 1 \times 10^{-13}$) by intein-Cas9(C574) than by wild-type Cas9. These findings indicate that specificity improvements of intein-Cas9 variants do not simply arise from differences in overall genome editing activity.
Intein 37R3-2 can be activated by other estrogen receptor modulators. To enable intein-Cas9 applications in which endogenous β-estradiol is present, we inserted into the estrogen receptor ligand-binding domain a point mutation (G521R) that renders the domain more specific for 4-HT. This mutation slightly reduces affinity for 4-HT but almost abolishes affinity for β-estradiol. The addition of this mutation to intein-Cas9(S219) eliminates the ability of β-estradiol to trigger Cas9 activity (Fig. 12).

The intein-Cas9 variants developed here demonstrate small-molecule control of Cas9 function, thereby enhancing genome-modification specificity. The use of ligand-dependent Cas9 variants provides greater control over genomic modification efficiencies and specificities than is currently achievable with constitutively active or transcriptionally regulated genome editing. This approach can synergize with other specificity-augmenting strategies such as direct delivery of transient Cas9 protein into cells, using truncated guide RNAs, paired Cas9 nickases, or FokI-ciCas9 fusions. This approach could also be applied to other genome engineering proteins to enable, for example, small-molecule control of TALE-based or Cas9-mediated transcriptional regulators.
EMX On  GAGTCCGAGCAGAAGAAGAAGGG (SEQ ID NO: XX)
EMX Off 1  GAGgCCGAGCAGAAGAAagACGG (SEQ ID NO: XX)
EMX Off 2  GAGTCtAGCAGgAGAAGAAGaG (SEQ ID NO: XX)
EMX Off 3  GAGTcTaAGCAGAAGAAGAGaG (SEQ ID NO: XX)
EMX Off 4  GAGTtAgGAGCAGAAGAAGaG (SEQ ID NO: XX)

VEGF On  GGGTGGGGGGGAGTTTGCTCCTGG (SEQ ID NO: XX)
VEGF Off 1  GGaTGGaGGGAGTTTGCTCCTGG (SEQ ID NO: XX)
VEGF Off 2  GGGaGGGtGGAGTTTGCTCCTGG (SEQ ID NO: XX)
VEGF Off 3  cGgGGaGGAGTTTGCTCCTGG (SEQ ID NO: XX)
VEGF Off 4  GGGgaGGGGaAGTTTGCTCCTGG (SEQ ID NO: XX)

CLTA On  GCAGATGTAGTGTTTCCACAGGG (SEQ ID NO: XX)
CLTA Off 1  aCAtATGTAGTaTTTCCACAGGG (SEQ ID NO: XX)
CLTA Off 2  cCAGATGTAGTaTTcCCACAGGG (SEQ ID NO: XX)
CLTA Off 3  ctAGATGaAGTGcTTCACATGG (SEQ ID NO: XX)

**Table 3.** On-target and 11 known off-target substrates of Cas9:sgRNAs that target sites in *EMX*, *VEGF*, and *CLTA*. List of genomic on-target and off-targets sites of the *EMX*, *VEGF*, and *CLTA* sites are shown with mutations from the on-target sequence shown in lower case. Protospacer-adjacent motifs (PAMs) are shown underlined.
<table>
<thead>
<tr>
<th></th>
<th>-4 H intein-Cas (S 19)</th>
<th>-4 H intein-Cas (C 74)</th>
<th>-4 H w Cas9 (50 ng)</th>
<th>+4-H intein-Cas (S 19)</th>
<th>+4-H intein-Cas (C 74)</th>
<th>+4-H w Cas (50 ng)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Indel</td>
<td>Tota</td>
<td>Mod Fre</td>
<td>Indel</td>
<td>Tota</td>
<td>Mod Fre</td>
</tr>
<tr>
<td>EM 0</td>
<td>12</td>
<td>597</td>
<td>1.7%</td>
<td>56</td>
<td>570</td>
<td>0.9%</td>
</tr>
<tr>
<td>EM Of 1</td>
<td>3</td>
<td>430</td>
<td>0.1%</td>
<td>4</td>
<td>354</td>
<td>0.1%</td>
</tr>
<tr>
<td>EM Of 2</td>
<td>8</td>
<td>532</td>
<td>0.2%</td>
<td>3</td>
<td>393</td>
<td>0.1%</td>
</tr>
<tr>
<td>EM Of 3</td>
<td>3</td>
<td>642</td>
<td>0.5%</td>
<td>1</td>
<td>452</td>
<td>0.2%</td>
</tr>
<tr>
<td>EM Of 4</td>
<td>14</td>
<td>763</td>
<td>0.9%</td>
<td>5</td>
<td>696</td>
<td>0.9%</td>
</tr>
<tr>
<td>VEG 0</td>
<td>35</td>
<td>309</td>
<td>1.5%</td>
<td>37</td>
<td>481</td>
<td>0.5%</td>
</tr>
<tr>
<td>VEG Of 1</td>
<td>21</td>
<td>433</td>
<td>0.3%</td>
<td>11</td>
<td>438</td>
<td>0.9%</td>
</tr>
<tr>
<td>VEG Of 2</td>
<td>2</td>
<td>352</td>
<td>0.8%</td>
<td>1</td>
<td>273</td>
<td>0.5%</td>
</tr>
<tr>
<td>VEG Of 3</td>
<td>1</td>
<td>464</td>
<td>0.4%</td>
<td>4</td>
<td>411</td>
<td>0.1%</td>
</tr>
<tr>
<td>VEG Of 4</td>
<td>5</td>
<td>572</td>
<td>0.0%</td>
<td>3</td>
<td>406</td>
<td>0.7%</td>
</tr>
<tr>
<td>CLT 0</td>
<td>28</td>
<td>456</td>
<td>4.0%</td>
<td>93</td>
<td>520</td>
<td>1.1%</td>
</tr>
<tr>
<td>CL Aor 1</td>
<td>8</td>
<td>708</td>
<td>0.1%</td>
<td>8</td>
<td>756</td>
<td>0.1%</td>
</tr>
<tr>
<td>CL Aor 2</td>
<td>3</td>
<td>613</td>
<td>0.0%</td>
<td>0</td>
<td>778</td>
<td>0.0%</td>
</tr>
<tr>
<td>CL Aor 3</td>
<td>1</td>
<td>432</td>
<td>0.0%</td>
<td>2</td>
<td>497</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

Table 4. Raw sequence counts and modification frequencies for data plotted in Figs. 4, 7, and 8. Total: total number of sequence counts. Modification frequency: number of indels divided by the total number of sequences listed as percentages.
<table>
<thead>
<tr>
<th></th>
<th>intein-Cas9(S219) (+ 4-HT vs. - 4-HT)</th>
<th>intein-Cas9(C574) (+ 4-HT vs. - 4-HT)</th>
<th>wt Cas9 (+ 4-HT vs. - 4-HT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMX On</td>
<td>&lt; 3.3 x 10^{-16}</td>
<td>&lt; 3.3 x 10^{-16}</td>
<td>1</td>
</tr>
<tr>
<td>VEGF On</td>
<td>&lt; 3.3 x 10^{-16}</td>
<td>&lt; 3.3 x 10^{-16}</td>
<td>1</td>
</tr>
<tr>
<td>CLTA On</td>
<td>&lt; 3.3 x 10^{-16}</td>
<td>&lt; 3.3 x 10^{-16}</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5. P-values for comparisons between conditions in Fig. 2a. P-values were obtained using the Fisher exact test and adjusted for multiple comparisons using the Benjamini-Hochberg Method.
Table 6. Raw sequence counts and modification frequencies for data plotted in Fig. 4b-d, and 9-11. Total: total number of sequence counts. Modification frequency: number of indels divided by the total number of sequences listed as percentages.
Table 7. P-values for comparisons between conditions in Fig. 7. P-values were obtained using the Fisher exact test and adjusted for multiple comparisons using the Benjamini-Hochberg Method.

<table>
<thead>
<tr>
<th></th>
<th>Independent Experiment 1</th>
<th>Independent Experiment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+ 4-HT intein- Cas9(S219) vs. + 4-HT intein- Cas9(C574)</td>
<td>+ 4-HT intein- Cas9(S219) vs. + 4-HT intein- Cas9(C574)</td>
</tr>
<tr>
<td></td>
<td>HT wt Cas9 (500 ng)</td>
<td>HT wt Cas9 (500 ng)</td>
</tr>
<tr>
<td>EMX On</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 2.4 x 10^-16</td>
</tr>
<tr>
<td>EMX Off 1</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 2.4 x 10^-16</td>
</tr>
<tr>
<td>EMX Off 2</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 3.9 x 10^-16</td>
</tr>
<tr>
<td>EMX Off 3</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 3.9 x 10^-16</td>
</tr>
<tr>
<td>EMX Off 4</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 3.9 x 10^-16</td>
</tr>
<tr>
<td>VEGF On</td>
<td>2.8 x 10^-12</td>
<td>&lt; 2.4 x 10^-16</td>
</tr>
<tr>
<td>VEGF Off 1</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 2.4 x 10^-16</td>
</tr>
<tr>
<td>VEGF Off 2</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 2.4 x 10^-16</td>
</tr>
<tr>
<td>VEGF Off 3</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 3.9 x 10^-16</td>
</tr>
<tr>
<td>VEGF Off 4</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 3.9 x 10^-16</td>
</tr>
<tr>
<td>CLTA On</td>
<td>1</td>
<td>&lt; 2.4 x 10^-16</td>
</tr>
<tr>
<td>CLTA Off 1</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 2.4 x 10^-16</td>
</tr>
<tr>
<td>CLTA Off 2</td>
<td>9.1 x 10^-5</td>
<td>4.4 x 10^-3</td>
</tr>
<tr>
<td>CLTA Off 3</td>
<td>1.3 x 10^-7</td>
<td>1.5 x 10^-14</td>
</tr>
</tbody>
</table>
Table 8. P-values for comparisons between conditions in Figs. 9 and 10. All conditions were treated with 4-HT. P-values were obtained using the Fisher exact test and adjusted for multiple comparisons using the Benjamini-Hochberg Method.

<table>
<thead>
<tr>
<th></th>
<th>intein C s9 S 19&lt;br&gt;vs w Cas9&lt;br&gt;(50 ng)</th>
<th>intein C s9 S 19&lt;br&gt;vs w Cas9&lt;br&gt;(26 ng)</th>
<th>intein C s9 S 19&lt;br&gt;vs w Cas9&lt;br&gt;(14 ng)</th>
<th>intein C s9 S 19&lt;br&gt;vs w Cas9&lt;br&gt;(5 ng)</th>
<th>intein C s9 C 74&lt;br&gt;vs w Cas9&lt;br&gt;(50 ng)</th>
<th>intein C s9 C 74&lt;br&gt;vs w Cas9&lt;br&gt;(26 ng)</th>
<th>intein C s9 C 74&lt;br&gt;vs w Cas9&lt;br&gt;(14 ng)</th>
<th>intein C s9 C 74&lt;br&gt;vs w Cas9&lt;br&gt;(8 ng)</th>
<th>intein C s9 C 74&lt;br&gt;vs w Cas9&lt;br&gt;(5 ng)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EM 0</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>1</td>
<td>1</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
</tr>
<tr>
<td>EM Of 1</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>1</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
</tr>
<tr>
<td>EM Of 2</td>
<td>&lt;3. x 1.3</td>
<td>4. x 10^-1</td>
<td>&lt;3. x 1.3</td>
<td>1</td>
<td>&lt;3. x 1.3</td>
<td>6. x 10^-1</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
</tr>
<tr>
<td>EM Of 3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>2. x 10^-1</td>
<td>&lt;3. x 1.3</td>
<td>2. x 10^-1</td>
<td>1. x 10^-1</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
</tr>
<tr>
<td>EM Of 4</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>1</td>
<td>&lt;3. x 1.3</td>
<td>3. x 10^-1</td>
<td>1. x 10^-1</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
</tr>
<tr>
<td>VEG 0</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>1</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
</tr>
<tr>
<td>VEG Of 1</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>3. x 10^-1</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
</tr>
<tr>
<td>VEG Of 2</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>5. x 10^-1</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
</tr>
<tr>
<td>VEG Of 3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>1. x 10^-1</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
</tr>
<tr>
<td>VEG Of 4</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>1</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
</tr>
<tr>
<td>CLT 0</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>1</td>
<td>1</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
</tr>
<tr>
<td>CLT Of 1</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>0.2</td>
<td>0.7</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
</tr>
<tr>
<td>CLT Of 2</td>
<td>1. x 10^-1</td>
<td>1. x 10^-1</td>
<td>0.1</td>
<td>9. x 10^-1</td>
<td>9. x 10^-1</td>
<td>9. x 10^-1</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
</tr>
<tr>
<td>CLT Of 3</td>
<td>3. x 10^-1</td>
<td>2. x 10^-1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.2</td>
<td>4. x 10^-1</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
<td>&lt;3. x 1.3</td>
</tr>
<tr>
<td>Indels</td>
<td>Total</td>
<td>Modification frequency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMX On</td>
<td>8</td>
<td>78943</td>
<td>0.01%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMX Off 1</td>
<td>1</td>
<td>42232</td>
<td>0.00%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMX Off 2</td>
<td>4</td>
<td>79008</td>
<td>0.01%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMX Off 3</td>
<td>60</td>
<td>113629</td>
<td>0.05%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMX Off 4</td>
<td>5</td>
<td>104159</td>
<td>0.00%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VEGF On</td>
<td>0</td>
<td>60667</td>
<td>0.00%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VEGF Off 1</td>
<td>2</td>
<td>111409</td>
<td>0.00%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VEGF Off 2</td>
<td>0</td>
<td>52048</td>
<td>0.00%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VEGF Off 3</td>
<td>4</td>
<td>88105</td>
<td>0.00%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VEGF Off 4</td>
<td>2</td>
<td>123559</td>
<td>0.00%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLTA On</td>
<td>491</td>
<td>68600</td>
<td>0.72%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLTA Off 1</td>
<td>10</td>
<td>116033</td>
<td>0.01%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLTA Off 2</td>
<td>6</td>
<td>75723</td>
<td>0.01%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLTA Off 3</td>
<td>4</td>
<td>53885</td>
<td>0.01%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 9. Raw sequence counts and modification frequencies (for cells transfected with wild-type Cas9 (500 ng) but without a targeting sgRNA, in the presence of 4-HT). Total: total number of sequence counts. Modification frequency: number of indels divided by the total number of sequences listed as percentages.
Sequences

[00151] Intein 37R32:

TGCCTTGCGAGGGTACCCGAATCTTCGATCCGGTCACTGGTACAACGCATCGCA
TCGAGGATGTTGTCGATGGGCGCAAGCCTATTCATGTCGTGGCTGCTGCCAAGGA
CGGAACGCTGCTCGCGCGGCCCGTGGTGTCCTGGTTCGACCAGGGAACGCGGGA
TGTGATCGGGTTGCGGATCGCCGGTGGCGCCATCGTGTGGGCGACACCCGATCAC
AAGGTGCTGACAGAGTACGGCTGGCGTGCCGCCGGGGAACTCCGCAAGGGAGAC
AGGGTGGCCGGACCGGGTGGTTCTGGTAACAGCCTGGCCTTGTCCCTGACGGCCG
ACCAGATGTTGGCTTGGTGTGACTGAGCCCCCATCTCATTTCGGAGTA
TGATCTACTAGCTCCCTCAGTGAAGCTCCCTGAGATGCTGACACCTCTTTTCTT
GGAGATCCTTCGACATGCTGCTGGCTACATCATCATTCGTTCCGAGCATATCTG
CAGGAGAGGTGTTGTTGCTGCTCAAATCTATTATTTGCTTATATTCTCAGATGT
ACACATTTCTGTCGGCGACACCTGGAATGCTTGGAGAGAAGACCATATCCACCG
AGCCCTGGACAAAGATGAGCAGACATTGATCCACTGCGCAGCCACAGCGCCT
GACCCTGACGCAGCAGCACCAGCGCTGCGCCAGCTCTCCTCTCATCCTCCTCAC
ATCAGGCACATGAGTAACAAAGAATGGGACATCAGTACAGTGAAGTACAAG
AACGTGGTGCCCCTCTATGACCTGCTGCTGGAGATGCTGGACGCCCACCGCCTAC
ATGCGGGTGGTTCTGGTGCTAGCCGCGTGCAGGCGTTCGCGGATGCCCTGGATGA
CAAATTCCTGACGACATGCTGCGGAGAGACTCCGCTATTCCGTGATCCGAGAA
GTGCTGCCAACCGCCGGGGCAGACGACGCTGCCCTCAGGAGGGAACCTGAC
ACCCTCGTCGCCGAAGGGGTTGTCGTGCACAACTGC

[00152] Cas9-NLS-3xFLAG:

MDKKYSIGLIDGINTNSVGAVITDEYKPSKKFVKLVGNTRDSIKKNLIGALLDFSGE
TAEATRLKRTARRYTRRKNRICLYQEIFSNEAKVDDSFFHRLLEESFVLVEEDKHE
RHPIFGNVDEVAHEYTAHYULRKLKVLDSTDKADLRLLYLALAHIMKFRGHLIEG
DLNPSDNSVDKLIQLQTVYNQLPENPAINASGVDAKAILAARLKLNLELAQLP
GEKKNGLFNGNLIAALSGLPTNFKSNFDLAEDAKLQLSILKLDYDDDDNLALLAIGDQA
YADFLLAAKLNSLDLISLQFADKLASMIRKYEHHQDLTKALVRQOLPE
KYEIEFFDQSKNYAGYIDGASQEYYYF KYF KPIKLEKMDGTEELVKLNRDLKLQKR
TFDNGSIPHIHLGELHJLRQQDFPAYFLKNREDREKILTFRPYVPVGLARGNRF
FMWMTKSEETTIPWNFEEVDDKGASAAQSFIERMTNFDKNLPNEKLPHSLLEYFVT
YNELTIIKVKIVTEGMRKPAFLSSEGQKKAIVDLEKTNKVRKVTKOLKYKTFICKCFD
SVEISVGREDFNASLGHTYDILKIIKDKDLNEDILEDILEITLTLFEDREMIEEL
KTYAHLFDKVMKQLKRRRTYGWRGRLSRKINGIRDQSKGTILDIFLSDGFANRN
FMQLIHHDSTLFKEDIQKAQVSQGYSQDSLHELHIANLPSAQKIKGILQTVKVDQ
VLMGRHKPENIVIEMARENQTQK,GQKNSRRERMKRIJEENIGELGQILKHEPVENTQL
QNEKLYLYLQNRGDMYVQDELRLSDYDVHIVPSLDFDSNKLVTRSDK
NRGKSNPVPEEYVKMKNYWQLNLAKLITQRKDFNLTAKERGSLDLEKAFI
RQLVRETQITKHVAQILDSSMNTK YEN DKL RSEV KVT LSKLSVDSFKDFQYKY
REINNYHHADYALVNVGTLAIKKEPKLYSEFQYDGKYVDKRVKMAKSEQEIKG
ATAKYFFYSXMANNNFKTEITLANGEIRKRPLETNGETGEIVWDKGRDFATVRKVL
SMQVNVKKTTEVQTFGSKESILPKNRSKDLIRKKDWDPKKYGGFDSTPVSAMLTVV
AKVEKGSKJKKSVEKELGITIMERSSSFKENPIDEALAGYKEVKKDLIKLPKYSFLE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIEQISEFSKRVLADANLDKVLASYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTIDRKRKYSTKEVLDAATLIHQSITGLYTERIDLSQLGGDGSPKKKRKVSSDYKDHDGDYKHDIDYKDDDDKAAG (SEQ ID NO: XX)

[00153] Intein-Cas9(C80)-NLS-3xFLAG:

MDKKYSIGLDDGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKLNIGALLFDSGE
TAEATRLKRTARRYTRRRNIECAGLAEGRIFDPVGTTHRIEDVDGRKPIHVVAACKDGTLLARPVVSFDQGTRVDIGLRIAGGAIWATPDHKVLTEYGWRAAGELRKGRVAGPGSSGNSLALSQTAQMSVALLDAEPPILYSEYDTPSFSEASMMGGLLTNLADRELVMHINAKRPGVFVDLTLDHQAHLERALWILEMIGLWVRSMEHPGKLLFAPLLNLDDNOGKCVEGVMVFQDLVRRFMRMMNQEEFVCVLSILLNSGVYFTLSSTLKSLEEKRAHDRTTDLHIHALMAKGLTLOOHOHRLALLILLSHIRMHMSNKGMEHLSMYKNVVPYVLDDLMDAHLRLHAGGSGARVVOAFADALDDKFLHMDLAEGLYSIVRLEPTPRTDFFLEEVEHTRLVEAVYHCYLEOFSEQSMNMAKVDSSFFHRLEESFLVEEDKHERPFGNIVIDVEAYHEKYPRYRKKLVDSTDKADRLILYLALARHMKRGHFLIGDLNPDVSLDQILQFVQTYNQLFEEPANGVDAKAILSARLSSKRSLNLAQPLGKEKGNGLNIALSLGLTNPFNKSDFLAEADKLQLSKTDYDLLLLNLAQIGDQYADFLAKNLDAILSDLIRPNEITEKAPLASMIRKYEDEHHDQTLKKALVRQOLPEIKYIFDSQSKNGYAGYIGDQASEQFYKFKKIPFMDGTEELLVKLREDLLRKQRTFDGSIPHIHLGELHAILRQEDFYPLKNDREKIEKLTFRPIPYVVPQPLAGNSRFAWMTRKSEEITTPWNEEVDKGAASOQFIERMNTDFKNLPNEKVLPKHSLLYEYFTVYNELTKVYETYEMGRPKAFLSEGKQKAIVDLLFFKTNRKVTVQKLKFEDYKFCCISVEIGVREDNFASLYIDHDLIIKIKDFDLNEENIDILEDIVLTLFLFEDREMIERLKYAHLDKDVKMQLARKRYTGWGBLSRKLINGIRDQSGKTILDFLKSDGFANFRFMQLLIHDDSLLTFKEIDIAQKAVSISGESGLSEHIANLAGSPAIIKKGILQTVKVEDLKVMGRHKPENIVIEMARENQTTQKQGNRSMRKRIEEGIKELGSQKILKEHPVENTQLQNEKLYLYYYQNQGRDMYDQEDLINSLDYDVIDHIVPSFLKDKDSIDNKVLTROSSRNKGSMDVPEEYVKKMKNYWQLLNAKLITQRFDNLTKAERGSLSDAKGFKIRQLVETFQITKHAVQLDSQSTNMKDENDKILEVKTITLKLVSDFRKFQFYKVERINNYHADLAYNLAVTGTAIIKYKPLESEFYVGYDKVYDVKRMIAQSEKDYAVKFYFNNMFICTEITLANIERKPLIETNEGETGEIVWDKGRDFATVRKLSMPQVNYKKTETVQGFSKEISLPKRNSDKLIARKKDWPKYYGGFDPSVTYAATVAVVEKGGSKKLKSVKELLGIMERSFSKPKFDFLEAGKYEVKKDLIIKLPKYSLEFLENGRLASKQNLGKSNLYVFNLAYSHYELKLPGGPEDEQFKQLFVQEIQHIKLYLDEIEQISEFSKRVLADANLDKVLASYNKHRDKPIREQAENIIHLFTLNLGAPAAFKYFDTIDRKRKYSTKEVLDAATLIHQSITGLYTERIDLSQLGGDGSPKKKRKVSSDYKDHDGDYKHDIDYKDDDDKAAG (SEQ ID NO: XX)

[00154] Intein-Cas9(A127)-NLS-3xFLAG:

MDKKYSIGLDDGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKLNIGALLFDSGETAEATRLKRTARRYTRRRNIECAGLAEGRIFDPVGTTHRIEDVDGRKPIHVVAACKDGTLLARPVVSFDQGTRVDIGLRIAGGAIWATPDHKVLTEYGWRAAGELRKGRVAGPGSSGNSLALSQTAQMSVALLDAEPPILYSEYDTPSFSEASMMGGLLTNLADRELVMHINAKRPGVFVDLTLDHQAHLERALWILEMIGLWVRSMEHPGKLLFAPLLNLDDNOGKCVEGVMVFQDLVRRFMRMMNQEEFVCVLSILLNSGVYFTLSSTLKSLEEKHIHALMAKGLTLOOHOHRLALLILLSHIRMHMSNKGMEHLSMYKNVVPYVLDDLMDAHLRLHAGGSGARVVOAFADALDDKFLHMDLAEGLYSIVRLEPTPRTDFFLEEVEHTRLVEAVYHCYLEOFSEQSMNMAKVDSSFFHRLEESFLVEEDKHERPFGNIVIDVEAYHEKYPRYRKKLVDSTDKADRLILYLALARHMKRGHFLIGDLNPDVSLDQILQFVQTYNQLFEEPANGVDAKAILSARLSSKRSLNLAQPLGKEKGNGLNIALSLGLTNPFNKSDFLAEADKLQLSKTDYDLLLLNLAQIGDQYADFLAKNLDAILSDLIRPNEITEKAPLASMIRKYEDEHHDQTLKKALVRQOLPEIKYIFDSQSKNGYAGYIGDQASEQFYKFKKIPFMDGTEELLVKLREDLLRKQRTFDGSIPHIHLGELHAILRQEDFYPLKNDREKIEKLTFRPIPYVVPQPLAGNSRFAWMTRKSEEITTPWNEEVDKGAASOQFIERMNTDFKNLPNEKVLPKHSLLYEYFTVYNELTKVYETYEMGRPKAFLSEGKQKAIVDLLFFKTNRKVTVQKLKFEDYKFCCISVEIGVREDNFASLYIDHDLIIKIKDFDLNEENIDILEDIVLTLFLFEDREMIERLKYAHLDKDVKMQLARKRYTGWGBLSRKLINGIRDQSGKTILDFLKSDGFANFRFMQLLIHDDSLLTFKEIDIAQKAVSISGESGLSEHIANLAGSPAIIKKGILQTVKVEDLKVMGRHKPENIVIEMARENQTTQKQGNRSMRKRIEEGIKELGSQKILKEHPVENTQLQNEKLYLYYYQNQGRDMYDQEDLINSLDYDVIDHIVPSFLKDKDSIDNKVLTROSSRNKGSMDVPEEYVKKMKNYWQLLNAKLITQRFDNLTKAERGSLSDAKGFKIRQLVETFQITKHAVQLDSQSTNMKDENDKILEVKTITLKLVSDFRKFQFYKVERINNYHADLAYNLAVTGTAIIKYKPLESEFYVGYDKVYDVKRMIAQSEKDYAVKFYFNNMFICTEITLANIERKPLIETNEGETGEIVWDKGRDFATVRKLSMPQVNYKKTETVQGFSKEISLPKRNSDKLIARKKDWPKYYGGFDPSVTYAATVAVVEKGGSKKLKSVKELLGIMERSFSKPKFDFLEAGKYEVKKDLIIKLPKYSLEFLENGRLASKQNLGKSNLYVFNLAYSHYELKLPGGPEDEQFKQLFVQEIQHIKLYLDEIEQISEFSKRVLADANLDKVLASYNKHRDKPIREQAENIIHLFTLNLGAPAAFKYFDTIDRKRKYSTKEVLDAATLIHQSITGLYTERIDLSQLGGDGSPKKKRKVSSDYKDHDGDYKHDIDYKDDDDKAAG (SEQ ID NO: XX)
Intein-Cas9(T146)-NLS-3xFLAG:

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKFFKVLGNTDRHSHIKKNNLIGALFDSGE TAEATRLKRTARRYTRKRNCRLCYQLQIEFSNEMAKVDDSFFHREESFLVEEDKHE RHPIFPFGNIVDEVAHYEKHTPIYRKLKVLDSCLAEGTRIFDPVTGTTTHREDVVDGRK PIHVVAAAKDGLLLRAPVVSFQDQTRVIGLRIAGAIWATDPHKVLTLEYGWRA AGELRKGDRVAAGPSSGSNLASLTQDQMYSVALLDEPPILYSEYDPTSPFSEASMM GLTNLADRELVHMINWASEPQVGVTLTLQDHAERLAMILGWVNRSMEH PKGLLFLAPNLLIDRNOGKCVGEVGMEIFDMLALLATSSRFMMNLQGEEVFVLCSIIIN SGLVYTFLLSTLSEELKMMNAKTMIDTILOMLHMACGIILTOOHLRAQLLLILSHI RHMSNSKGMELHYSMLKNVPPYDLLEMDLAHRLHAGGSGARVOAFADALDD KFLHDLAEGLRSPVLPTRARFDVEEVLHTLVAEGVYHVNCYHEKYPTIYRKLKVLVSDTADLR LIY LALAHMIFRQHGFLIEGIDLNPDSVDKLFLIQVIQTQNYQLFEEVINPASGVDAKAIL ARSLKSRRLENLIAQLPGKEKNGLFLGNIALSLGLTPNKSFDLADEAKLQSLKTDY DDDLDNLLAQIGDQYADLFRAAKLSDLIDLRVNTEITKAPLAMSIKMKYDEHH HQDTLKALVRQQLPEKYEIFFDQSKNGYAGYIDGGAQSEQFHYFKFPIKLEMGD TEELLKVLNREDLQRTFDSNPISIQHILGHAIRLRRQEDFYLFKLDNEKIEKIL TFRIPYYVGPLARGNSRFMTRKEETITTPWNFEEVDGKASAQSFIERMTNDKDN LNPEVKLHSSLLYEFTVYNETLKVYVTENGMRPFSLSEQGKAKIAVLDLKLFTNKR VTVQKLKEDYFKKIECFDSVEISVGDREFNALSGTYHDLKIIKDDFSLDEENEDILE DIVTLLTFEDREMIEERLKYTAHLFDDKVQKLRRRTGWGRLSKINGIRDQK SGKTIDLFKSDLGAFNRMFLHDSLTFKEDIQKAKQVSGQDSLHEHIANLAGSPA IKKGILQTQTVKVLKVMRGRHPIVENIMARENQTQTKQGKNSRERMKRIEGIK ELGSLQIKHEPVENTQLNEKLYLYQLONGDMYQDVLNRSDYDVHIVPQSF LKDDSINVKLTRSDKRNGSDKNVSEEVKKMKNYWRQLLNAKILITQRFSDKNLTK AERGGLSLDKAGFIKRQLVETRQITKHVAQILDSRMNYTDENKLRLEKVVTILKS KLVSDFRKDFQYFKVREINNYHHAHDYLANVGTALIKKYKLESEFYVDGYKVY DVKRMIAKSEQEIGKATAKYSFNSMFFKITEQLANGERKPLIETNGETEIVWD KGRDFATVRKLSMPQNVINIKTTEGVTFGQSESKLPSNKDLRIKARDDWPKKY GGFDSTPVTAVSLLVVAKEGKSKKLLSKVEKLGGITMERSSEFFKNPIDFLEAKGYKE VKKDLIIKLPKYSFLEENRGLKMLAEALQEGKALPSKYNFLYSLHYEKLK GSPDEQKQLFLVEQHIIYLDIEIEQISEFSKRVILADNLKLSAYNKHRDKPIRE QAENIIHFLTLPGAPAAFYKDTRKYSTKVLDTIALHIQSTLGYETRIDLQ LGGDGSPKKKRKVSDDYKDHGDGYDKDDIDYKDDDDKAAG (SEQ ID NO: XX)
AERGGLSELDKAGFIKRQLVETRQITKHVAQILDSTMNTKYDENDKLI</p>
IVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKCAVQSQGDSLHEHIANLAGSPA
I KKGILQTVKVDLEVKLVMGRHKPENIVIEMARENQTTQKGKNSRERMKRIEEGIK
ELGSQILKEHPVENTIQNEKELYLYQNGRMYVQDELINRLSDYVDHIVPSQSF
LKDSSDNKVLTRSDKNGKSNDVPNSEEVVKMKNYWRQLLNAKLITQRKFDNLTK
AERGGLSELDKAGFIKQLVETRQITKVQILDSRMNTKYDENKDLIREVKVTILKS
KLVSDFKRFKDYFYKREVINYHYHADAYLNAVGTALIKKYKLESEFYVGYKY
DVRKMIAKSEIEGKATAKYYFYSNMFFKTEILTOLANGEIRKPLIETNTGETEIVWD
KGRDFATVRKLSPMQVNIKKETEVQTFGFSKESILPKRNSDKLIAKRDWPKYY
GGFDSTVAYSVQVAKVEGGSKKLKSYKELLGITIMERSSEKFNPDIFLEAKGYKE
VKKDIIKLPKSYLFELENCRKLASAGELQKGNELAPLSKYNFLYSLYSHYKLK
GSPDENQKQLFVEQKHELEFEIESFKVRILADNLDKLVSAYKNKRDKPIRE
QAENIIHLFLTLNLGAPAAFKYFDTTIDKRKYTSTKEVLGATLHIQSITGLYETRIDSQ
LGGDGSPKKRRKVVSSDYKDHDGDYKDHDIDYKDDDDKAA (SEQ ID NO: XX)
Intein-Cas9(T622)-NLS-3xFLAG:

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSE
TAEATRLKRTARRYTRRKNRICYLQEIFSNEAMAVDSFFHRLEESFLVEEDKKHE
RHPFIGNVDEVAYHEKYPTIYHRLKVKLDSTDKADRLYIALAHMIKFRGHLIEG
DLNPDNSVDKLFIQLQVTQNYQLFEENPINASVDAAKAILSARLKSRRLENLIAQLP
GEKKNGLFGNLLASLGLTNPKLQFDNYLFADAKQLQSKDTRYDDELDNLAQIGDQYA
DLFLAANKLSDAILSDLERNITEKAPLSASMIKRYDEHHQDLTLLKALVRQOLPE
KYKEIFFDQSQKNGYAGIDGASQEEFYKFKPILEKMDGTTEELLVKLREDLLRKRQ
TFDNSIPIHQIHLGELHAILRRQEDFYPFLKDNREKIEFLTRFYVGPLRNSRFA
WMTRKSEETrPWRFQEEVDKASQSFIERMTNFDKLNEVKLPKSLLYEFYTV
YNELTKVYTEDGMKPAMLSEGQKIAVDLFLKRTKVQKLKEDFYKIECIFD
SVEISGVEDRFNASLGYTDDLKIKDKDLDNENEDILEDIVLCLAEGTRIFDVPVTG
THHIEDVVDGRKPIHVHFAADGTLALLARPVVSFDQGTRDVGLRIAGGAIVWATP
DHKVLTEYGWRAAGELRKGDRVAGPGGSGNSLAFLTADQMSALLDAEPPILYSE
YDPTSPFSEAMSMLFTNLADRELVHMINWAKRVPFGVDTLTHQAHILRERALEWLEI
LMIGLWVRSMHEHGPKLIQDNLDRNOGKCVEGMVEIFDLMLATSSRFRMMNLQ
GEEFVCLKSIILLNSGYTFITLSSTLKSLEEKDHIHRALDKITDTHLMAKAGTLLOQ
HORLQALLLILQEMHLYKSVKYNVPVLDELLMDAHRLHAGSSG
ASRVOAFADALDDDDHLMELGRYSVIREPLTRARTFDLEVELHTLVAEGV
VHVClFTLVEEQMLEERKTYAHLFDDVKKMVKLRTYGVWGRSLKLINGIRDK
QSGKTILDFKSDGFANRNFQMQFLSDLTFFKEDIQAQVSQGSGDSLHEHIANLGSP
AIKKGILQTVKVVDELVKVMGRHKPENIVIEARENQTTQKGKQSNRKRKIEGII
KELQSGKIKHPVNTQLQNEKLKYLYLQNRKMYVQDELQIRLSYVDHIVQPV
SFLLKSHSQVTRSDRNKGRSDNJSVEPEWVKMKNYWQRNLAKRITQDFNL
TKAEGGGLSELQDKAFIRQRVETRIKTHVALDLSRMTKYENDKLIREVKVITL
KSKLVDSDKFDQFQYKVREINNHYHAHADYNLANVTGLYKPKLESEFYDYKV
VYDVRKMIAKSEQEIGKATAKYFFYSIMIMNFKTEITLANEGIRKPLIIETNGETGEIV
WDKGRDFATVRKLVMPQVNVKKEVQVTQGFSKESLPKNSDKLKAIKWDWPDK
KYGGFDSPTVAYSVLVLAVKVEKGKSSSKLKVKEGTMKERSFENPQDFLEAKG
YKVEKKDLIIKLPYKSLFELENGKRMALASQGCNGALPSLYKYNFLLYASHYE
KLKGSPEDNEQKQLFVEQKHKYLDEIEIQISEFSKRVIADANLDKLVSAYNKHDRDK
IREQAENIIHLFTTLNLAGAPAEAFYKPDTTDTRKRKTSTKVEDATLHIQSnrGLYETRID
LSQLLGGDSKPKKRRVKVSYDHDGKDHYKDDDKAAG (SEQ ID NO: XX)

Intein-Cas9(S701)-NLS-3xFLAG:

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSE
TAEATRLKRTARRYTRRKNRICYLQEIFSNEAMAVDSFFHRLEESFLVEEDKKHE
RHPFIGNVDEVAYHEKYPTIYHRLKVKLDSTDKADRLYIALAHMIKFRGHLIEG
DLNPDNSVDKLFIQLQVTQNYQLFEENPINASVDAAKAILSARLKSRRLENLIAQLP
GEKKNGLFGNLLASLGLTNPKLQFDNYLFADAKQLQSKDTRYDDELDNLAQIGDQYA
DLFLAANKLSDAILSDLERNITEKAPLSASMIKRYDEHHQDLTLLKALVRQOLPE
KYKEIFFDQSQKNGYAGIDGASQEEFYKFKPILEKMDGTTEELLVKLREDLLRKRQ
TFDNSIPIHQIHLGELHAILRRQEDFYPFLKDNREKIEFLTRFYVGPLRNSRFA
WMTRKSEETrPWRFQEEVDKASQSFIERMTNFDKLNEVKLPKSLLYEFYTV
YNELTKVYTEDGMKPAMLSEGQKIAVDLFLKRTKVQKLKEDFYKIECIFD
SVEISGVEDRFNASLGYTDDLKIKDKDLDNENEDILEDIVLCLAEGTRIFDVPVTG
THHIEDVVDGRKPIHVHFAADGTLALLARPVVSFDQGTRDVGLRIAGGAIVWATP
DHKVLTEYGWRAAGELRKGDRVAGPGGSGNSLAFLTADQMSALLDAEPPILYSE
YDPTSPFSEAMSMLFTNLADRELVHMINWAKRVPFGVDTLTHQAHILRERALEWLEI
LMIGLWVRSMHEHGPKLIQDNLDRNOGKCVEGMVEIFDLMLATSSRFRMMNLQ
GEEFVCLKSIILLNSGYTFITLSSTLKSLEEKDHIHRALDKITDTHLMAKAGTLLOQ
HORLQALLLILQEMHLYKSVKYNVPVLDELLMDAHRLHAGSSG
ASRVOAFADALDDDDHLMELGRYSVIREPLTRARTFDLEVELHTLVAEGV
VHVClFTLVEEQMLEERKTYAHLFDDVKKMVKLRTYGVWGRSLKLINGIRDK
QSGKTILDFKSDGFANRNFQMQFLSDLTFFKEDIQAQVSQGSGDSLHEHIANLGSP
AIKKGILQTVKVVDELVKVMGRHKPENIVIEARENQTTQKGKQSNRKRKIEGII
KELQSGKIKHPVNTQLQNEKLKYLYLQNRKMYVQDELQIRLSYVDHIVQPV
SFLLKSHSQVTRSDRNKGRSDNJSVEPEWVKMKNYWQRNLAKRITQDFNL
TKAEGGGLSELQDKAFIRQRVETRIKTHVALDLSRMTKYENDKLIREVKVITL
KSKLVDSDKFDQFQYKVREINNHYHAHADYNLANVTGLYKPKLESEFYDYKV
VYDVRKMIAKSEQEIGKATAKYFFYSIMIMNFKTEITLANEGIRKPLIIETNGETGEIV
WDKGRDFATVRKLVMPQVNVKKEVQVTQGFSKESLPKNSDKLKAIKWDWPDK
KYGGFDSPTVAYSVLVLAVKVEKGKSSSKLKVKEGTMKERSFENPQDFLEAKG
YKVEKKDLIIKLPYKSLFELENGKRMALASQGCNGALPSLYKYNFLLYASHYE
KLKGSPEDNEQKQLFVEQKHKYLDEIEIQISEFSKRVIADANLDKLVSAYNKHDRDK
IREQAENIIHLFTTLNLAGAPAEAFYKPDTTDTRKRKTSTKVEDATLHIQSnrGLYETRID
LSQLLGGDSKPKKRRVKVSYDHDGKDHYKDDDKAAG (SEQ ID NO: XX)
Intein-Cas9(A728)-NLS-3xFLAG:

MDKKYSIGLDIGTNSVGWAVITDEYEVKSPKKFKVLGNTDRHSKKNLIGALLFDSGE
TAEATRLKRYTRRKYTRRKNRICTQLEIFSNEAMAKVDDFHRILVEEDKKE
RHPIFGNVDEVAHEKypiHTKLRKKLVDSTKDADILYLALAHMIFRGHFLIEG
DLNPDVSVDKLFIQILVTYQTINQLFENPINASVDAKALSALSRLSKRNLLEAQLP
GEKKGFLGNLIALSLSDKNFDLAEADKLQLSKDCTDDLLNALLAQIGDQYA
DLFRAAKNLSDAIIISDLIRNTEIPLASMIKRYDEHHQDLTLLKALVRQOLPE
KYEIIFDQQSNKYAGYIGDGQSGINYEFKPIEKMDGTEELVKLNREDLRKRQ
TFDNGSIPHIHLGELHAIILRQEDYFYPFLKDNREIKTIFRIPYVVPGLARNSRFA
WMTRKSEETITPWNEEEVKDGASASQFIERMTNFDKLPNKELPLLEYEFTV
YNELTIVKYTEGMRKPAFLSGEQKKAIVDLFKRNDQTVKQKLEDYKFIKECD
SVEISGVEDRFNASLGYHTDILKIJKDFDLNENEDILEDIVLTLTFEDREMIEERL
KTYAHLLDFKVMKQLKRRYTGWGRLSKLIRGKDQSGKTILDFKSLDSGFAANR
FMOLIHDDSDIKLIOIQAQVSGOSGLSHEHIANLCLAEGRIFDPVCNTTIRLEDV
DGRIHIIVAAAKDGTLLARPVVVSWDQGRVTRDVGLRIAGGAIWVATPDHVKLTYE
GWRAAGELRGDRVAGPGSNSGLSLSLTAQDSMVALLDEAPPILYDPTPSFSE
ASMMGLTNTLADRELHMINWAKRPGFVDLTHDAHLLEARWEILMIGLVWR
SMехалгfiKLFFAPNLLLDRNOKGVCVEG
MEIFDMATASSRMMNQLGEEFVCLKSIILLNSGVYFLSSTLKSLEEKDHIIHRA
LDKITDTLIIMAKAGLTOOHORLAIQLL
ILSHERHMNSKMEGKYKVNPILGDYKIDDDKKAAG (SEQ ID NO: XX)

[00162]
Intein-Cas9(T995)-NLS-3xFLAG:
MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKNKLIGALLFDGSE TAAEATRLKRTARRYTRRRKNIRCYLQEIFSNEMAKVDSFFHRLEESFLVEEDKKHE RHPFIGNVDEAYHEKYPTIYHLRLKVLVSTDKDACLRLIYLALAHMIKFRGHHFLIEG DLPDNDSVDKLFIQLQVTQNYLQNEPINAAGVDALKISLRSKRLLENIALQLP GEKKNGLFNLALSGLTLPNFKSNFDLAEDAKLQLSĐTYYDDDLNLAIQIDQYA DLFLAANKLSDAILSLDIFLRNVEITKAPLASSMIRYDEHHQDLTLKLAVQOLPKE KYIEIFDQSKNGYAGYIDGGASEQEEFYKFIKPILEKMDGTEEELVKLNREDLRKQR TFDNGSIPHIQLHELHAILLRQQEDFYPFLKDNREKIEKLITFRIPYYVGPLARGNSRFA WMTRKSEETrPWNFEEVVDGBKASAOFSFIERMNTDFKDLNPEKVLPSHSLLEYEYFTV YNELTKVKYYTEMKPAFLSSEGQKKAIVDLFFKTNRKVTVQKLEDYFKKIECFD SVEISVGQEAFLNSALGTYHDLLKIIKDDFLNEEDILEDIVTLTLFEDREIELRKT KHYAHLFDDKVMGRWYGRSLRLRIRDRQGSKTDLKDSGFANNR FMQLIHDDSLSTKEDIQAQSQGQGDSLHEHIANLAGSPAIIKQILGQTVKVVDLVK VMGRHKPENIVIEMRQENHTQGKQKNSRERKMREEGIEIKGLSQILKEHPVENTQL QNEKLYLYLQNGRMDYVQDELNRSDLQDDYDVHIQPSFVLDDSDKDSIKNLVTRSDK NRGKSDNVSEELGTLYRQLNATLKRIFDLNTKAEGRGLSELDKAQFIK RQLVETRQITKHVAILDSLRSMTKYYENDKLREVKVTLKLVDKRDFQFYKV REINNYHHHAYDHAYLNAVCGAEGRTRIFDPVTGTTHRIEDVDGRKPIHVVAAA KD GTLLARPVYFWHDIGLLAGAIVGAVWPDHKVLTEYWGRAAGELRKGRD VRAPPGSSGNLALSITADQMRVSAALDEAPILSETYDPSFSEASMMGLTNLADR ELVHMNWKAPVFVDTLISDAMHLLAWEILMGTVSREHPGKLAPNL LDDRNOGKVEGMVEIFDMLATSSRFMNLQGEEVFCLKSILLNSGYTFLSTL KSLEEKDIHRLDKITDITLHIMAKALTLLOOHHORLALAQLLLLHHRHMSMKGME HLHYSMKYNVVPYLIDLLMADHLRLHAGGSGASROAFADALDDKFLHDMLAE GLRYVSIREVLTPTRRATFDFLEVEEHTLVAEIGVWVHNACLIKKYPKLESEFYVGYD KVYDVRKMIAKSEQEIGKATAYYFYSEMSNFKFETLANGEIRKPRIETNETGEI VWDKGRDFATVRKVLMSPQVNIIVKKTEVQTTGFSKESIPKRRNSDKLIARKKDWDPK KKYGFFDSPVTAYSVLVVAKEVGKSKKLVSKEVGLLITTIMRESSFSENPIDLEAK GYVEKVKDLIIKPKYSLFELENGRKMLASAGELQKGENALPSKVYNFLLSASHY EKLKGSPIPENOKQLFVEQHKHYLDEIEQISEFSKRVILADANLDKVLASYNKHREDK PIREQAEINIIHFTLNLGAAPAAYKFFDYMontDKRITYSTKEVAILATLHISQITLYETR DLSQLQGGGDGSPPKKRRKVSSDYKDHGDYKDHIDYKDDDDKAAG (SEQ ID NO: XX)

Intein-Cas9(S1006)-NLS-3xFLAG:
MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKNKLIGALLFDGSE TAAEATRLKRTARRYTRRRKNIRCYLQEIFSNEMAKVDSFFHRLEESFLVEEDKKHE RHPFIGNVDEAYHEKYPTIYHLRLKVLVSTDKDACLRLIYLALAHMIKFRGHHFLIEG DLPDNDSVDKLFIQLQVTQNYLQNEPINAAGVDALKISLRSKRLLENIALQLP GEKKNGLFNLALSGLTLPNFKSNFDLAEDAKLQLSĐTYYDDDLNLAIQIDQYA DLFLAANKLSDAILSLDIFLRNVEITKAPLASSMIRYDEHHQDLTLKLAVQOLPKE KYIEIFDQSKNGYAGYIDGGASEQEEFYKFIKPILEKMDGTEEELVKLNREDLRKQR TFDNGSIPHIQLHELHAILLRQQEDFYPFLKDNREKIEKLITFRIPYYVGPLARGNSRFA WMTRKSEETrPWNFEEVVDGBKASAOFSFIERMNTDFKDLNPEKVLPSHSLLEYEYFTV YNELTKVKYYTEMKPAFLSSEGQKKAIVDLFFKTNRKVTVQOLKEDYFKKIECFD
Intein-Cas9(Sl 154)-NLS-3xFLAG:

MDKKYSGLDGTNSVGWAVITDEYKVPSKKKFVLGNTDHSIKKNNLIGALLLDSGE
TAEATRLKRTARRRYTRRKRNCYQEFSN EMAKVDSSHFRLLESFVLEDKKHE
RHPIFPGNVDVEAYHEKPTYIYHLKRLV DSTKANLRLYILALAHMKFRGHFLIG
DLNPDDSVD KFLIQVLSQYNEPISAGVDK ALSRLSKRR LNAQP L
GEKKNLGFLNIALSLGLTPNSFKSNFDLAEADKQLSKDTSDDLNLQA IQGDQ Y
DLFLAANKNLSAIII DILR VNEIT YKAPL SASM I K M YKDHEQDHLT KALVRQQLPE
KYIEIFFDQSKNGYAGYIDGGA S QEE FYKF IKPKLE KMDGTEELVLKLNREDLLRQK
TFDNGSIPHQLHELHAIRLRRQEFYPDPLKNDREKIEKLTRFYVGPLAR NSFA
WMTRKSKETrPWNFEVVDGKASAQSFIERN TDFKDNLPKSH VLE YFTV
YNELTKVY VTEGRFSQEGKQAI VDLF NTFRKTVQK LE FYK KIECFD
SVEISGVEDRNASL GLTYHD LKIIDKD FDNEEN DILEDIVLTL TLFDREMIEERL
KTYAHLLFKDVKMQ LKKRRYTG WRLSRLKINGRDKQG TIDLFLKSDGFANRN
FMQLIHDSSLTKEDIQKAQVSGQGDSLHEHIANLAGSPA IKKG ILQTVKVD ELVK
VMGHRKPEVIENAMENRQTTQGQKNR MKEEEGKIGLSQLKEH PVENTQL
QNEKLVLYLQ LNGRMDYVQDELINRSL DVFHIVPQSFKL DSDNKVLTRSDK
NRGSKDNPSEEVVKMMKNYWQRLNAMLQESTKR DKFNLTKAERG SED L DKA FIK
RQLVETRQITKHVA QILDNRMNTK YDENKLEIR KTVL KSLSVDSRKFDFQFYKV
REINLYHAHDAYLNAVGTALIKPKLESEVF YDYVDRKMI ASEIELGK ATKAYFFYSNMFFKETEITLANGEIRKRPI L TGETGEIVWDKGRDFATVRK VLSM
PQVNIKVTQE QTGGSKNASRKDNSDLKAI RKWDPPKYYGFDSPTVAS YLV
AKVEKGKCLAEGRIFDPVGT TTHRIEDVGBKPIHVVA AAKDGTLARSEA

SEQ ID NO: XX}

WO 2016/022363

PCT/US2015/04770

110
PLYDLLLEMLDAHRLHAGGSGARVOAFADALDDKFLHDLMAEGLRYSVIREVLPTRRARTFDLLEEHLTLVAEGVVVHVNCCKKLKSKVELLGITIMERSSFEKPNIDPLEAEGKYEVEKLDLIQLPKYSLFLENGRKRMLASAGELQKNELALPSKYVYNFLYLASHYEKLKGSPEDNEQFVLQKYHLFIYDIQESFRKVLICAFANDLKDVLASAYNKHRDKPIREQAENIIHLFTLNLGAPAAAFKYFDTTIDRKRYPTSTKEVLDATLIHQSSrGLYETRIDSQQLGDDGSPKKKRKKVSSDYKDHGDGYKDHIDYKDDDDKAAG (SEQ ID NO: XX)

[00166] Intein-Cas9(S1159)-NLS-3xFLAG:
MDKKYSIGLIGTNSVGWAVITDEYKVPSSKFKVGLGNTDRHISIKNKLIGALLFDSGE
TAEATRLKRTARRYTRRRKNRCYLQEIFSNEAMAVIDDSFFHRLEESFLVEEDKKHE
RHPIFQNVDEVAYHEKPYHTLRLKVLVDSTKDALRLYIALAHMIKFRHFLIEG
DLPNPDSDVDKLFQILQVQTYNLQFEENPASVGDAKAILSARLKSRRLENLIAQLP
GEKKNGLFNLASLGLGTPNFKSNFDLADNLKQSLKDTTYDDDLNAQIDQYA
DLFLAANLSDLASLILRVNTEITKPLSASMIKRYDEHQQDLTLKLALVRQQLPE
KYKIEFFQDSQNAGYIDGASQEEFYFIPKILEKMDGTEELLVKLNDDELRKRQRTFNGSPHQHLGELHARQLEDDFFQKLNDQRIEIIKLELTFRPPYVGPLARGNSFAWMTKRKSEETrrPWNFFEEVDKGASAQSFIERMRTNDKNPLNEKLVPLKSHLLJETFTVYNELTKVYKVPMLAFLSEQGKKAIVDLLFKTNKVRTVQKLDYESYFKKIECD
SVEISVGEDFNASLGTMYHDLKFLKDNLNEEDEDILVTLTLFDREMEIERL
KTYAHLFFDVMQVKLKQRYTGWGLRSLKINGRQDSQKGLTIDFLSDKGAFANRN
FMQLIHDDSSTFKEDIQKAQVSGQGDSLHEHIANLASSPAIKKGLQITKVVDDELK
VMGRHKPENIVIEMARENQTQTQKQKNSRMRKMRIEGIKGLCEGQLKEHPVENTQL
QNEKLYYLQLNQGRMDSYLVDQSVPLKDSIDNKLVRSTRS
NRGKSDNPSVEVYKKMNYYRQNLAKLITQRKFDNLTKAQERGLSLEDKAGFK
RQVLQVETRQKVQILDSRMTNKTYDNEKLKREVQVLTQSKLSVDFKRDFOQKYV
REINNYHAHADYLANAVTDLKJYKPELESEFYGYKDYVRMIKIASEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNETGEIVWDKGRDFATVRKLSM
PQVNIVKNTEVTQGFSKESILEPKNRSKDIALKRDWDPPYGGFSDPVTAYSLLV
AKVEKKGSKKKLKCLAEGTRIFDPPVTGTTTHRIEDYVDGRKPHVVAADGTLARPV
VSWFDPQGTRDVIGLRAGAIJWATPDHKLTVETYGWAAELKRGDRVAGPGGSNSLALSXQMSVALLDAEPILYSEYDTPSFSEASMMGLLTNLADRELVHMINW
AKRVPFGVDLTHDHQAHLERALWLEIMQILMVRSMEHPSGKLLFAPNLLDDRNOKGCVEGMVEIFDMDLATSSRFMRMLNQGEFVCLKSIILSundayNQTVFLSSTSLSLEKDH
IHRLADKTDITLHIHMALKATLNOOHORLALQLILLIIHRHMSNKMGHELMSMKYA
NVVPYDDLLLEMLDAHRLHAGGSGARVOFADFADALDDKFLHDLMAEGLRYSVIRE
VLPTRARTFDLLEEHLTLVAEGVVVHVNCCKKLKSKVELLGITIMERSSFEKPNIDPLEAEGKYEVEKLDLIQLPKYSLFLENGRKRMLASAGELQKNELALPSKYVYNFLYLASHYEKLKGSPEDNEQFVLQKYHLFIYDIQESFRKVLICAFANDLKDVLASAYNKHRDKPIREQAENIIHLFTLNLGAPAAAFKYFDTTIDRKRYPTSTKEVLDATLIHQSSrGLYETRIDsQQLGDDGSPKKKRKKVSSDYKDHGDGYKDHIDYKDDDDKAAG (SEQ ID NO: FX)

[00167] Intein-Cas9(S1274)-NLS-3xFLAG:
MDKKYSIGLIGTNSVGWAVITDEYKVPSSKFKVGLGNTDRHISIKNKLIGALLFDSGE
TAEATRLKRTARRYTRRRKNRCYLQEIFSNEAMAVIDDSFFHRLEESFLVEEDKKHE
RHPIFQNVDEVAYHEKPYHTLRLKVLVDSTKDALRLYIALAHMIKFRHFLIEG
DLPNPDSDVDKLFQILQVQTYNLQFEENPASVGDAKAILSARLKSRRLENLIAQLP
GEKKNGLFNLASLGLGTPNFKSNFDLADNLKQSLKDTTYDDDLNAQIDQYA
DLFLAANLSDLASLILRVNTEITKPLSASMIKRYDEHQQDLTLKLALVRQQLPE
KYKIEFFQDSQNAGYIDGASQEEFYFIPKILEKMDGTEELLVKLNDDELRKRQRTFNGSPHQHLGELHARQLEDDFFQKLNDQRIEIIKLELTFRPPYVGPLARGNSFAWMTKRKSEETrrPWNFFEEVDKGASAQSFIERMRTNDKNPLNEKLVPLKSHLLJETFTVYNELTKVYKVPMLAFLSEQGKKAIVDLLFKTNKVRTVQKLDYESYFKKIECD
SVEISVGEDFNASLGTMYHDLKFLKDNLNEEDEDILVTLTLFDREMEIERL
KTYAHLFFDVMQVKLKQRYTGWGLRSLKINGRQDSQKGLTIDFLSDKGAFANRN
FMQLIHDDSSTFKEDIQKAQVSGQGDSLHEHIANLASSPAIKKGLQITKVVDDELK
VMGRHKPENIVIEMARENQTQTQKQKNSRMRKMRIEGIKGLCEGQLKEHPVENTQL
QNEKLYYLQLNQGRMDSYLVDQSVPLKDSIDNKLVRSTRS
NRGKSDNPSVEVYKKMNYYRQNLAKLITQRKFDNLTKAQERGLSLEDKAGFK
RQVLQVETRQKVQILDSRMTNKTYDNEKLKREVQVLTQSKLSVDFKRDFOQKYV
REINNYHAHADYLANAVTDLKJYKPELESEFYGYKDYVRMIKIASEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNETGEIVWDKGRDFATVRKLSM
PQVNIVKNTEVTQGFSKESILEPKNRSKDIALKRDWDPPYGGFSDPVTAYSLLV
AKVEKKGSKKKLKCLAEGTRIFDPPVTGTTTHRIEDYVDGRKPHVVAADGTLARPV
VSWFDPQGTRDVIGLRAGAIJWATPDHKLTVETYGWAAELKRGDRVAGPGGSNSLALSXQMSVALLDAEPILYSEYDTPSFSEASMMGLLTNLADRELVHMINW
AKRVPFGVDLTHDHQAHLERALWLEIMQILMVRSMEHPSGKLLFAPNLLDDRNOKGCVEGMVEIFDMDLATSSRFMRMLNQGEFVCLKSIILSundayNQTVFLSSTSLSLEKDH
IHRLADKTDITLHIHMALKATLNOOHORLALQLILLIIHRHMSNKMGHELMSMKYA
NVVPYDDLLLEMLDAHRLHAGGSGARVOFADFADALDDKFLHDLMAEGLRYSVIRE
VLPTRARTFDLLEEHLTLVAEGVVVHVNCCKKLKSKVELLGITIMERSSFEKPNIDPLEAEGKYEVEKLDLIQLPKYSLFLENGRKRMLASAGELQKNELALPSKYVYNFLYLASHYEKLKGSPEDNEQFVLQKYHLFIYDIQESFRKVLICAFANDLKDVLASAYNKHRDKPIREQAENIIHLFTLNLGAPAAAFKYFDTTIDRKRYPTSTKEVLDATLIHQSSrGLYETRIDLSQQLGDDGSPKKKRKKVSSDYKDHGDGYKDHIDYKDDDDKAAG (SEQ ID NO: XX)
Intein-Cas9(S219-G^21R)-NLS-3xFLAG:

MDKKYSIGLDIGTNVSNGWAVITDEYKVPSKKFKVLGLNTDHRISIKNNLIGALLFDSGE
TAAEATRKLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYTIYHLRRKVLVDSTDKADLRLLIYLALAHMIKFRGHFLIESG
DLNPDNSVDKLFILOVTVONLFEENPINASVGDAKAILRSKLCLAEITRIFDPVTGV
TTHIRDVYVRGKPIHVVAADKTLLARPVVSFDGOTDRVGLRIAGGAVWL
ATPDHKVLEYGRAERLGDRAVAGPGGSGNLASLTDAMQVSAILDAPPIL
YSEYDPTSPFSEASMMGLLNLADRELVHMINWAKRVPGFVDDLTLHDQAHELLRA
WEILMLGVRSSMEHPKLLFDNRSDLNDRGOKCVAEGMVFEDLMATSSRFRMM
NLQGEEFVCKSIIILLNSGVYVFTLSSTLKSLEEKDIHRLDKITDDHLMAMAGLTL
LOOHORLAQLILLSHRHSNKJRMEHYLSMKYKVNVPDLYDLLEMADLRHAG
GSGARVOAFADALDDLKDHFMDLEAELRSYIREVLPLTRARTFDELEEHTLVAE
GVVVHNCRLNIELVACRGKNSNGLSALSLGTLPNFKSNFDLAEADKLQSKD
TYDDDLNLIQDGYYAFLDLAANLSDAILLSDILRVTNEITKAPLSAMIKRYDE
HHQDLTLKALVRQPLKYEIIFFDSKNGYAGIDGASQEEFYKFIKPILEKMD
GTEELVKNREDLRLKQRTTFDNSISPISIQHLGELHAIIRRQEDFYPFLKDRENIEK
LTRFRPPYGPGARNSFWAMTRKSEETrtPWNEEEVDGKASAQSFIERMTNFDK
NLPNEKLVPHKSLLEYFTVYNELTKVITYEGRMKPAFLSSEQKKAIVDLFLKTNR
KVTVKLQKEDYFKKICFDSVEISVGVEDRFNASGLTYHDLKIIKDKDFLDNEENEDI
LEDIVTTLTLEDREMIERLKYAHLDSDKVMKLRRYGTVWRGLSRKLLINGIRD
KQSGKTLDFKSDGFANRFMVHIHDSTFTKIIEDKIAKVSGQGDLSHELHIANL
SPAIIKKGILQTVKVDDELVKMGRHPHENIVIEMARENQTQKQGKNSRERMKRIE

112
Indel Calling Algorithm

Read 1:

@M00265:68:000000000-AA85W: 1:1101: 14923: 1642 1:N:0: 1
TGCA
TCTCATGACTTGGCCTTTGTAGGAAAACACCATTAGAAGAG TAGATGGTT GGGTAGTGGCTCCTCTCTGGCTTTAGACTCTTTGGCTACTATGAAATAAAGGCTCTTAT

TTGCAAAAGGCGGTGTATGGGTTGAAGCACATTGAGAAAGAGGCT

Read 2:

@M00265:68:000000000-AA85W: 1:1101: 14923: 1642 2:N:0: 1
CTCACCTGGGCGAGAAAGGTAACTTATGTTTCAGTAGCCTCTTTCTCAATGTGC

CAACCCATACCGGCTTTTGCAATAGAG CCCTTTATCCATAGTAGACAA GAGTCT AAGCAGAAGAGAGCCACTACCCAACCATCTACTCTCTATGGT

Step 1: Search for sequences (or reverse complements) flanking the on/off target sites in both Illumina reads from the following set:

<table>
<thead>
<tr>
<th>target site</th>
<th>5' flanking sequence</th>
<th>3' flanking sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMX_On</td>
<td>GAGTCCGAGCAAGAAGAGGG</td>
<td>AGCTGGAGGAGGAGGGGCCT</td>
</tr>
<tr>
<td>EMX_Off1</td>
<td>GAGGCCGAGCAAGAAGAGGC</td>
<td>CCCCTTCTCTGAAATGAG</td>
</tr>
<tr>
<td>EMX_Off2</td>
<td>GAGTCCGAGCAAGAAGAGGC</td>
<td>GGCTGGGGGCAGCATGACCT</td>
</tr>
<tr>
<td>EMX_Off3</td>
<td>GAGTCTAGCAGAAGAAGAGGC</td>
<td>GACAGGATGTTGGGGGAG</td>
</tr>
<tr>
<td>EMX_Off4</td>
<td>GAGTGGAGGAGGAGGAGAGGC</td>
<td>CATGGCAAGAGATGTCAG</td>
</tr>
</tbody>
</table>
Step 2: Extract the sequence between the target sites in both reads and ensure that it is identical (reverse complementary) in read 1 and read 2 and all positions within read 1 and read 2 have a quality score $\geq 30$ (Phred score $\geq 30$)

In above reads, CTCTTCTGCTTAGACTC is reverse complement of GAGTCTAAGCAGAAGAG

Step 3: Align extracted sequence to the reference sequence for the relevant on/off target sequence

GAGTCTAAGCAGAAGAG reference sequence
GAGTCTAAGCAGAAGAG sequence read

Step 4: For deletions, count only if deletion occurred in close proximity to expected cleavage site (within 8 bp of 3’ end of reference sequence)

Methods and Materials

[00169] Cas9, intein-Cas9, and sgRNA expression plasmids. A plasmid encoding the human codon-optimized Streptococcus pyogenes Cas9 nuclease with an NLS and 3xFLAG tag (Addgene plasmid 43861)\(^5\) was used as the wild-type Cas9 expression plasmid.

Intein 37R3-2 was subcloned at the described positions into the wild-type Cas9 expression plasmid using USER (NEB M5505) cloning. sgRNA expression plasmids used in this study have been described previously\(^1\)\(^6\). Plasmid constructs generated in this work will be deposited with Addgene.

[00170] Modification of genomic GFP. HEK293-GFP stable cells (GenTarget), which constitutively express Emerald GFP, served as the reporter cell line. Cells were maintained in "full serum media": Dulbecco's Modified Eagle's Media plus GlutaMax (Life
Technologies) with 10% (vol/vol) FBS and penicillin/streptomycin (lx, Amresco). 5 x 10⁴ cells were plated on 48-well collagen-coated Biocoat plates (Becton Dickinson). 16-18 h after plating, cells were transfected with Lipofectamine 2000 (Life Technologies) according to the manufacturer's protocol. Briefly, 1.5 µg of Lipofectamine 2000 was used to transfect 650 ng of total plasmid: 500 ng Cas9 expression plasmid, 125 ng sgRNA expression plasmid, and 25 ng near-infrared iRFP670 expressing plasmid (Addgene plasmid 45457).

12 h after transfection, the media was replaced with full serum media, with or without 4-HT (1 µM, Sigma-Aldrich T176). The media was replaced again 3-4 days after transfection. Five days after transfection, cells were trypsinized and resuspended in full serum media and analyzed on a C6 flow cytometer (Accuri) with a 488-nm laser excitation and 520-nm filter with a 20-nm band pass. Transfections and flow cytometry measurements were performed in triplicate.

**[00171]** High-throughput DNA sequencing of genome modifications. HEK293-GFP stable cells were transfected with plasmids expressing Cas9 (500 ng) and sgRNA (125 ng) as described above. For treatments in which a reduced amount of wild-type Cas9 expression plasmid was transfected, pUC19 plasmid was used to bring the total amount of plasmid to 500 ng.

4-HT (1 µM final), where appropriate, was added during transfection. 12 h after transfection, the media was replaced with full serum media without 4-HT. Genomic DNA was isolated and pooled from three biological replicates 60 h after transfection using a previously reported protocol with a DNAdvance Kit (Agencourt). 150 ng or 200 ng of genomic DNA was used as a template to amplify by PCR the on-target and off-target genomic sites with flanking HTS primer pairs described previously. PCR products were purified using RapidTips (Diffinity Genomics) and quantified using the PicoGreen dsDNA Assay Kit (Invitrogen). Purified DNA was PCR amplified with primers containing sequencing adaptors, purified with the MinElute PCR Purification Kit (Qiagen) and AMPure XP PCR Purification (Agencourt). Samples were sequenced on a MiSeq high-throughput DNA sequencer (Illumina), and sequencing data was analyzed as described previously.

**[00172]** Western blot analysis of intein splicing. HEK293-GFP stable cells were transfected with 500 ng Cas9 expression plasmid and 125 ng sgRNA expression plasmid. 12 h after transfection, the media was replaced with full serum media, with or without 4-HT (1 µM). Cells were lysed and pooled from three technical replicates 4, 8, 12, or 24 h after 4-HT treatment. Samples were run on a Bolt 4-12% Bis-Tris gel (Life Technologies). An anti-FLAG antibody (Sigma-Aldrich F1804) and an anti-mouse 800CW IRDye (LI-COR) were used to visualize the gel on an Odyssey IR imager.
[00173] **Statistical analysis.** Statistical tests were performed as described in the figure captions. All p-values were calculated with the R software package. p-values for the Fisher exact test were calculated using the fisher.test function, with a one-sided alternative hypothesis (alternative = "greater" or alternative = "less", as appropriate). Upper bounds on p-values that are close to zero were determined manually. The Benjamini-Hochberg adjustment was performed using the R function p.adjust (method = "fdr").

[00174] **Sensitivity limit of off-target cleavage assays.** We used paired end sequencing to identify indels caused by genomic on- and off-target cleavage. Given that published studies (see the reference below) have shown that the Illumina platform has an indel rate that is several orders of magnitude lower than the -0.1% substitution error rate, and our requirement that all called indels occur in both paired reads, the sensitivity of the high-throughput sequencing method for detecting genomic off-target cleavage in our study is limited by the amount genomic DNA (gDNA) input into the PCR amplification of each genomic target site. A 1 ng sample of human gDNA represents only -330 unique genomes, and thus only -330 unique copies of each genomic site are present. PCR amplification for each genomic target was performed on a total of 150 ng or 200 ng of input gDNA, which provides amplicons derived from at most 50,000 or 65,000 unique gDNA copies, respectively. Therefore, the high-throughput sequencing assay cannot detect rare genome modification events that occur at a frequency of less than approximately 1 in 50,000 (0.002%). When comparing between two conditions, such as wt Cas9 vs. intein-Cas9, this threshold becomes approximately 10 in 50,000 (0.02%) when using the Fisher exact test and a conservative multiple comparison correction (Bonferroni with 14 samples). See also Minoche, A. E., Dohm, J. C, & Himmelbauer, H. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. *Genome Biology* 12, R112 (2011).

**REFERENCES**


All publications, patents, patent applications, publication, and database entries (e.g., sequence database entries) mentioned herein, e.g., in the Background, Summary, Detailed Description, Examples, and References sections, are incorporated by reference in their entirety as if each individual publication, patent, patent application, publication, and database entry was specifically and individually incorporated herein by reference. In case of conflict, the present application, including any definitions herein, will control.

**EQUIVALENTS AND SCOPE**

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. The scope of the present invention is not intended to be limited to the above description, but rather is as set forth in the appended claims.

In the claims articles such as "a," "an," and "the" may mean one or more than one unless indicated to the contrary or otherwise evident from the context. Claims or descriptions that include "or" between one or more members of a group are considered satisfied if one, more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process unless indicated to the contrary or otherwise evident from the context. The invention includes embodiments in which exactly one member of the group is present in, employed in, or otherwise relevant to a given product or process. The invention also includes embodiments in which more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process.

Furthermore, it is to be understood that the invention encompasses all variations, combinations, and permutations in which one or more limitations, elements, clauses, descriptive terms, etc., from one or more of the claims or from relevant portions of the description is introduced into another claim. For example, any claim that is dependent on another claim can be modified to include one or more limitations found in any other claim that is dependent on the same base claim. Furthermore, where the claims recite a composition, it is to be understood that methods of using the composition for any of the
purposes disclosed herein are included, and methods of making the composition according to any of the methods of making disclosed herein or other methods known in the art are included, unless otherwise indicated or unless it would be evident to one of ordinary skill in the art that a contradiction or inconsistency would arise.

[00179] Where elements are presented as lists, e.g., in Markush group format, it is to be understood that each subgroup of the elements is also disclosed, and any element(s) can be removed from the group. It is also noted that the term "comprising" is intended to be open and permits the inclusion of additional elements or steps. It should be understood that, in general, where the invention, or aspects of the invention, is/are referred to as comprising particular elements, features, steps, etc., certain embodiments of the invention or aspects of the invention consist, or consist essentially of, such elements, features, steps, etc. For purposes of simplicity those embodiments have not been specifically set forth in haec verba herein. Thus for each embodiment of the invention that comprises one or more elements, features, steps, etc., the invention also provides embodiments that consist or consist essentially of those elements, features, steps, etc.

[00180] Where ranges are given, endpoints are included. Furthermore, it is to be understood that unless otherwise indicated or otherwise evident from the context and/or the understanding of one of ordinary skill in the art, values that are expressed as ranges can assume any specific value within the stated ranges in different embodiments of the invention, to the tenth of the unit of the lower limit of the range, unless the context clearly dictates otherwise. It is also to be understood that unless otherwise indicated or otherwise evident from the context and/or the understanding of one of ordinary skill in the art, values expressed as ranges can assume any subrange within the given range, wherein the endpoints of the subrange are expressed to the same degree of accuracy as the tenth of the unit of the lower limit of the range.

[00181] In addition, it is to be understood that any particular embodiment of the present invention may be explicitly excluded from any one or more of the claims. Where ranges are given, any value within the range may explicitly be excluded from any one or more of the claims. Any embodiment, element, feature, application, or aspect of the compositions and/or methods of the invention, can be excluded from any one or more claims. For purposes of brevity, all of the embodiments in which one or more elements, features, purposes, or aspects is excluded are not set forth explicitly herein.

[00182] All publications, patents and sequence database entries mentioned herein, including those items listed above, are hereby incorporated by reference in their entirety as if
each individual publication or patent was specifically and individually indicated to be incorporated by reference. In case of conflict, the present application, including any definitions herein, will control.
CAS9 PROTEINS INCLUDING LIGAND-DEPENDENT INTEINS

RELATED APPLICATIONS


GOVERNMENT SUPPORT

[0002] This invention was made with U.S. Government support under R01 GM095501 and F32GM 106601, awarded by the National Institutes of Health/National Institute of General Medical Sciences, and under grant numbers HR001 1-1-2-0003 and N66001-12-C-4207, awarded by the Department of Defense. The Government has certain rights in the invention.

BACKGROUND OF THE INVENTION

[0003] Site-specific enzymes theoretically allow for the targeted manipulation of a single site within a genome and are useful in the context of gene targeting as well as for therapeutic applications. In a variety of organisms, including mammals, site-specific enzymes such as endonucleases have been used for genome engineering by stimulating either non-homologous end joining or homologous recombination. In addition to providing powerful research tools, site-specific nucleases also have potential as gene therapy agents, and two site-specific endonucleases have recently entered clinical trials: one, CCR5-2246, targeting a human CCR-5 allele as part of an anti-HIV therapeutic approach (clinical trials NCT00842634, NCT01044654, NCT01252641), and the other one, VF24684, targeting the human VEGF-A promoter as part of an anti-cancer therapeutic approach (clinical trial NCT01082926).

[0004] Specific manipulation of the intended target site without or with only minimal off-target activity is a prerequisite for clinical applications of site-specific enzymes, and also for high-efficiency genomic manipulations in basic research applications. For example, imperfect specificity of engineered site-specific binding domains of certain nucleases has been linked to cellular toxicity and undesired alterations of genomic loci other than the intended target. Most nucleases available today, however, exhibit significant off-target
activity, and thus may not be suitable for clinical applications. An emerging nuclease platform for use in clinical and research settings are the RNA-guided nucleases, such as Cas9. While these nucleases are able to bind guide RNAs (gRNAs) that direct cleavage of specific target sites, off-target activity is still observed for certain Cas9:gRNA complexes (Pattanayak et al., "High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity." Nat Biotechnol. 2013; doi: 10.1038/nbt.2673). Technology for engineering site-specific enzymes with reduced off-target effects is therefore needed.

SUMMARY OF THE INVENTION

The reported toxicity of some engineered site-specific enzymes such as endonucleases is thought to be based on off-target DNA cleavage. Further, the activity of existing RNA-guided nucleases generally cannot be controlled at the molecular level, for example, to switch a nuclease from an "off" to an "on" state. Controlling the activity of nucleases and other site-specific enzymes suitable for nucleic acid manipulations or modifications could decrease the likelihood of incurring off-target effects. Some aspects of this disclosure provide strategies, compositions, systems, and methods to control the binding and/or enzymatic activity of RNA-programmable enzymes, such as Cas9 endonuclease, nickases, deaminases, recombinases, transcriptional activators and repressors, epigenetic modifiers variants and fusions thereof.

Accordingly, one aspect of the present disclosure provides Cas9 proteins (including fusions of Cas9 proteins and functional domains) comprising inteins, for example, ligand-dependent inteins. The presence of the intein inhibits one or more activities of the Cas9 proteins, for example, nucleic acid binding activity (e.g., target nucleic acid binding activity and/or gRNA binding activity), a nuclease activity, or another enzymatic activity (e.g., nucleic acid modifying activity, transcriptional activation and repression, etc.) for which the Cas9 protein (e.g., Cas9 fusion protein) is engineered to undertake (e.g., nuclease activity, nickase activity, recombinase activity, deaminase activity, transcriptional activator/repressor activity, epigenetic modification, etc.). In some embodiments, the Cas9 protein is a Cas9 nickase. The Cas9 fusions are typically between a nuclease inactivated Cas9 ("dCas") and one or more functional domains. The intein may be inserted into any location of a Cas9 protein, including one or more domains of a Cas9 protein or Cas9 fusion (including in a functional domain), such as the HNH nuclease domain or the RuvC nuclease domain. In some embodiments, the intein replaces amino acid residue Cys80, Ala127, Thr146, Ser219, Thr333, Thr519, Cys574, Thr622, Ser701, Ala728, Thr995, Ser1006,
Ser1 154, Serl 159, or Serl274 in the Cas9 polypeptide sequence set forth as SEQ ID NO:2, in
the dCas9 polypeptide sequence set forth as SEQ ID NO:5, or in the Cas9 nickase
polypeptide sequence set forth as SEQ ID NO:4. In some embodiments, the intein replaces
or is inserted at an amino acid residue that is within 5, within 10, within 15, or within 20
amino acid residues of Cys80, Ala27, Thr146, Ser219, Thr333, Thr519, Cys574, Thr622,
Ser701, Ala728, Thr995, Serl006, Serl 154, Serl 159, or Serl274 in the Cas9 polypeptide
sequence set forth as SEQ ID NO:2, in the dCas9 polypeptide sequence set forth as SEQ ID
NO:5, or in the Cas9 nickase polypeptide sequence set forth as SEQ ID NO:4. the intein
replaces amino acid residue Ala27, Thr146, Ser219, Thr519, or Cys574 in the Cas9
polypeptide sequence set forth as SEQ ID NO:2, in the dCas9 polypeptide sequence set forth
as SEQ ID NO:5, or in the Cas9 nickase polypeptide sequence set forth as SEQ ID NO:4.
Typically the intein is a ligand-dependent intein which exhibits no or minimal protein
splicing activity in the absence of ligand (e.g., small molecules such as 4-hydroxytamoxifen,
peptides, proteins, polynucleotides, amino acids, and nucleotides). Ligand-dependent inteins
are known, and include those described in U.S. patent application, U.S.S.N. 14/004,280,
published as U.S. 2014/0065711 Al, the entire contents of which are incorporated herein by
reference. In some embodiments, the intein comprises an amino acid sequence selected from
the group consisting of SEQ ID NO:7-14.

[0007] In one aspect, a Cas9 protein is provided that comprises: (i) a nuclease-
inactivated Cas9 (e.g., dCas9 (SEQ ID NO:5)) domain; (ii) a ligand-dependent intein; and
(iii) a recombinase catalytic domain. In some embodiments, the ligand-dependent intein
domain is inserted into the dCas9 domain as described herein. Typically, the presence of the
intein in the Cas9 protein inhibits one or more activities of the Cas9 protein, such as gRNA
binding activity, target nucleic acid binding activity, and/or recombinase activity.
Accordingly, upon self-excision of the intein (e.g., induced by ligand binding the intein) the
one or more activities of the Cas9 protein is/are restored. In some embodiments, the
recombinase catalytic domain is a monomer of the recombinase catalytic domain of Hin
recombinase, Gin recombinase, or Tn3 recombinase.

[0008] According to another aspect, a Cas9 protein is provided that comprises: (i) a
nuclease-inactivated Cas9 (e.g., dCas9 (SEQ ID NO:5)) domain; (ii) a ligand-dependent
intein; and (iii) a deaminase catalytic domain. In some embodiments, the ligand-dependent
intein domain is inserted into the dCas9 domain as described herein. Typically, the presence
of the intein in the Cas9 protein inhibits one or more activities of the Cas9 protein, such as
gRNA binding activity, target nucleic acid binding activity, and/or deaminase activity.
Accordingly, upon self-excision of the intein (e.g., induced by ligand binding of the intein) the one or more activities of the Cas9 protein is/are restored. In some embodiments, the deaminase catalytic domain comprises a cytidine deaminase (e.g., of apolipoprotein B niRNA-editing complex (APOBEC) family deaminases such as APOBEC1 or activation-induced cytidine deaminase (AID)). In some embodiments, the deaminase catalytic domain comprises a ACF1/ASE deaminase or an adenosine deaminase, such as a ADAT family deaminase.

[0009] Accordingly, upon self-excision of the intein (e.g., induced by ligand binding of the intein) the one or more activities of the Cas9 protein is/are restored. In some embodiments, the epigenetic modifier domain is selected from the group consisting of protein, such as gRNA binding activity, target nucleic acid binding activity, and/or transcriptional activation. Accordingly, upon self-excision of the intein (e.g., induced by ligand binding the intein) the one or more activities of the Cas9 protein is/are restored. In some embodiments, the transcriptional activator domain is VP64, CP16, and p65.

[0010] According to yet another aspect, a Cas9 protein is provided that comprises: (i) a nuclease-inactivated Cas9 (e.g., dCas9 (SEQ ID NO:5)) domain; (ii) a ligand-dependent intein; and (iii) a transcriptional activator domain. In some embodiments, the ligand-dependent intein domain is inserted into the dCas9 domain as described herein. Typically, the presence of the intein in the Cas9 protein inhibits one or more activities of the Cas9 protein, such as gRNA binding activity, target nucleic acid binding activity, and/or transcriptional activation. Accordingly, upon self-excision of the intein (e.g., induced by ligand binding the intein) the one or more activities of the Cas9 protein is/are restored. In some embodiments, the transcriptional activator domain is KRAB, SID, or SID4x.

According to yet another aspect, a Cas9 protein is provided that comprises: (i) a nuclease-inactivated Cas9 (e.g., dCas9 (SEQ ID NO:5)) domain; (ii) a ligand-dependent intein; and (iii) an epigenetic modifier domain. In some embodiments, the ligand-dependent intein domain is inserted into the dCas9 domain as described herein. Typically, the presence of the intein in the Cas9 protein inhibits one or more activities of the Cas9 protein, such as gRNA binding activity, target nucleic acid binding activity, and/or epigenetic modification activity. Accordingly, upon self-excision of the intein (e.g., induced by ligand binding the intein) the one or more activities of the Cas9 protein is/are restored. In some embodiments, the epigenetic modifier domain is epigenetic modifier is selected from the group consisting of...
histone demethylase, histone methyltransferase, hydroxylase, histone deacetylase, and histone acetyltransferase. In some embodiments, the epigenetic modifier comprises the LSD1 histone demethylase or TET1 hydroxylase.

[0011] According to another aspect, methods of using Cas9 proteins are provided. In some embodiments involving site-specific DNA cleavage, the methods comprise (a) contacting a Cas9 protein (e.g., having nuclease activity) comprising a ligand-dependent intein with a ligand, wherein binding of the ligand to the intein induces self-excision of the intein; and (b) contacting a DNA with the Cas9 protein, wherein the Cas9 protein is associated with a gRNA; whereby self-excision of the intein from the Cas9 protein in step (a) allows the Cas9 protein to cleave the DNA, thereby producing cleaved DNA. In some embodiments, the Cas9 protein first binds a gRNA and optionally the target DNA prior to excision of the intein, but is unable to cleave the DNA until excision of the intein occurs. Any of the Cas9 proteins having nuclease activity and comprising a ligand-dependent intein, as described herein, can be used in the inventive methods.

[0012] According to another aspect, methods of using any of the ligand-dependent intein-containing Cas9 proteins comprising a recombinase catalytic domain are provided. In some embodiments, the method is useful for recombinating two nucleic acids, such as two DNAs, and comprises (a) contacting a first DNA with a first ligand-dependent dCas9-recombinase fusion protein (e.g., any of those described herein), wherein the dCas9 domain of the first fusion protein binds a first gRNA that hybridizes to a region of the first DNA; (b) contacting the first DNA with a second ligand-dependent dCas9-recombinase fusion protein, wherein the dCas9 domain of the second fusion protein binds a second gRNA that hybridizes to a second region of the first DNA; (c) contacting a second DNA with a third ligand-dependent dCas9-recombinase fusion protein, wherein the dCas9 domain of the third fusion protein binds a third gRNA that hybridizes to a region of the second DNA; and (d) contacting the second DNA with a fourth ligand-dependent dCas9-recombinase fusion protein, wherein the dCas9 domain of the fourth fusion protein binds a fourth gRNA that hybridizes to a second region of the second DNA; whereby the binding of the fusion proteins in steps (a) - (d) results in the tetramerization of the recombinase catalytic domains of the fusion proteins, under conditions such that the DNAs are recombined. In some embodiments, the methods are useful for site-specific recombinating between two regions of a single DNA molecule, and comprise (a) contacting the DNA with a first ligand-dependent dCas9-recombinase fusion protein, wherein the dCas9 domain if the first fusion protein binds a first gRNA that hybridizes to a region of the DNA; (b) contacting the DNA with a second ligand-dependent
dCas9-recombinase fusion protein, wherein the dCas9 domain of the second fusion protein binds a second gRNA that hybridizes to a second region of the DNA; (c) contacting the DNA with a third ligand-dependent dCas9-recombinase fusion protein, wherein the dCas9 domain of the third fusion protein binds a third gRNA that hybridizes to a third region of the DNA; (d) contacting the DNA with a fourth ligand-dependent dCas9-recombinase fusion protein, wherein the dCas9 domain of the fourth fusion protein binds a fourth gRNA that hybridizes to a fourth region of the DNA; whereby the binding of the fusion proteins in steps (a) - (d) results in the tetramerization of the recombinase catalytic domains of the fusion proteins, under conditions such that the DNA is recombined. In some embodiment, any of the methods first comprise contacting the fusion proteins with a ligand that induces self-excision of the intein. In some embodiments, the fusion proteins are contacted with the ligand after: (i) the fusion proteins bind a gRNA; (ii) the fusion proteins bind the DNA; or (iii) after the recombinase domains form a tetramer. In some embodiments, the gRNAs in any step (a)-(d) of the inventive methods hybridize to the same strand or to opposing strands in the DNA(s). In some embodiments, the gRNAs hybridize to regions of their respective DNAs that are no more than 10, no more than 15, no more than 20, no more than 25, no more than 30, no more than 40, no more than 50, no more than 60, no more than 70, no more than 80, no more than 90, or no more than 100 base pairs apart.

According to yet another aspect, methods of using any of the ligand-dependent intein Cas9 proteins comprising deaminase catalytic domains are provided. The methods comprise contacting a DNA molecule with (a) a ligand-dependent Cas9 protein comprising deaminase catalytic domain as provided herein; and (b) a gRNA targeting the Cas9 protein of step (a) to a target nucleotide sequence of the DNA strand; wherein the DNA molecule is contacted with the Cas9 protein, and the gRNA in an amount effective and under conditions suitable for the deamination of a nucleotide base. In some embodiments, the methods comprise contacting the Cas9 protein with a ligand that induces self-excision of the intein either before or after the Cas9 protein binds the gRNA. In some embodiments, the target DNA sequence comprises a sequence associated with a disease or disorder, and wherein the deamination of the nucleotide base results in a sequence that is not associated with a disease or disorder. In some embodiments, the DNA sequence to be modified comprises a T→C or A→G point mutation associated with a disease or disorder, and the deamination of the mutant C or G base results in a sequence that is not associated with a disease or disorder (e.g., the deamination corrects the mutation the caused the disease or disorder). In some embodiments, the deamination corrects a point mutation in the sequence associated with the disease or
disorder. In some embodiments, the sequence associated with the disease or disorder encodes a protein, and wherein the deamination introduces a stop codon into the sequence associated with the disease or disorder, resulting in a truncation of the encoded protein. In some embodiments, the deamination corrects a point mutation in the PI3KCA gene, thus correcting an H1047R and/or a A3140G mutation. In some embodiments, the contacting is performed in vivo in a subject susceptible to having or diagnosed with the disease or disorder. In some embodiments, the disease or disorder is a disease associated with a point mutation, or a single-base mutation, in the genome. In some embodiments, the disease is a genetic disease, a cancer, a metabolic disease, or a lysosomal storage disease.

[0014] According to another aspect, methods for transcriptional activation of a gene are provided. In some embodiments, the methods comprise contacting a DNA molecule comprising a gene with (a) a ligand-dependent dCas9 fusion protein comprising a transcriptional activator (e.g., any of those provided herein) and (b) a gRNA targeting the fusion protein of (a) to a target nucleotide sequence of the DNA strand; wherein the DNA molecule is contacted with the fusion protein and the gRNA in an amount effective and under conditions suitable for the transcriptional activation of the gene. In some embodiments, the method further comprises contacting the fusion protein with a ligand that induces self-excision of the intein. In some embodiments, the fusion protein is contacted with the ligand prior to forming a complex with a gRNA. In some embodiments, the fusion protein is contacted with the ligand after forming a complex with a gRNA. In some embodiments, the gRNA targets the promoter region of a gene.

[0015] According to another aspect, methods for transcriptional repression of a gene are provided. In some embodiments, the methods comprise contacting a DNA molecule comprising a gene with (a) a ligand-dependent dCas9 fusion protein comprising a transcriptional repressor (e.g., any of those provided herein) and (b) a gRNA targeting the fusion protein of (a) to a target nucleotide sequence of the DNA strand; wherein the DNA molecule is contacted with the fusion protein and the gRNA in an amount effective and under conditions suitable for the transcriptional repression of the gene. In some embodiments, the method further comprises contacting the fusion protein with a ligand that induces self-excision of the intein. In some embodiments, the fusion protein is contacted with the ligand prior to forming a complex with a gRNA. In some embodiments, the fusion protein is contacted with the ligand after forming a complex with a gRNA. In some embodiments, the gRNA targets the promoter region of a gene.
According to another aspect, methods for epigenetic modification of DNA are provided. In some embodiments, the DNA is chromosomal DNA. In some embodiments, the methods comprise contacting a DNA molecule with (a) a ligand-dependent dCas9 fusion protein comprising a epigenetic modifier (e.g., any of those provided herein) and (b) a gRNA targeting the fusion protein of (a) to a target nucleotide sequence of the DNA strand; wherein the DNA molecule is contacted with the fusion protein and the gRNA in an amount effective and under conditions suitable for the epigenetic modification of the DNA. In some embodiments, the method further comprises contacting the fusion protein with a ligand that induces self-excision of the intein. In some embodiments, the fusion protein is contacted with the ligand prior to forming a complex with a gRNA. In some embodiments, the fusion protein is contacted with the ligand after forming a complex with a gRNA. In some embodiments, the gRNA targets the promoter region of a gene in the DNA.

Any of the methods provided herein can be performed on DNA in a cell, for example, a cell in vitro or in vivo. In some embodiments, any of the methods provided herein are performed on DNA in a eukaryotic cell. In some embodiments, the eukaryotic cell is in an individual, for example, a human.

According to some embodiments, polynucleotides are provided, for example, that encode any of the proteins (e.g., proteins comprising ligand-dependent Cas9 proteins or variants) described herein. In some embodiments, vectors that comprise a polynucleotide described herein are provided. In some embodiments, vectors for recombinant expression of any of the proteins (e.g., comprising ligand-dependent Cas9 proteins or variants) described herein are provided. In some embodiments, cells comprising genetic constructs for expressing any of the proteins (e.g., comprising ligand-dependent Cas9 proteins or variants) described herein are provided.

In some embodiments, kits useful in using, producing, or creating any of the ligand-dependent Cas9 proteins or variants thereof, as described herein, are provided. For example, kits comprising any of the proteins (e.g., ligand-dependent Cas9 proteins or variants) described herein are provided. In some embodiments, kits comprising any of the polynucleotides described herein are provided. In some embodiments, kits comprising a vector for recombinant expression, wherein the vectors comprise a polynucleotide encoding any of the proteins (e.g., ligand-dependent Cas9 proteins or variants) described herein, are provided. In some embodiments, kits comprising a cell comprising genetic constructs for expressing any of the proteins (e.g., ligand-dependent Cas9 proteins or variants) described herein are provided.
Other advantages, features, and uses of the invention will be apparent from the Detailed Description of Certain Embodiments of the Invention; the Drawings, which are schematic and not intended to be drawn to scale; and the Claims.

**BRIEF DESCRIPTION OF THE DRAWINGS**

Figure 1 shows a schematic depicting an exemplary embodiment of the disclosure. A Cas9 protein comprising a ligand-dependent intein, remains inactive in the absence of a ligand that binds the intein domain. Upon addition of the ligand, the intein is self-excised, restoring the activity of the Cas9 protein. Cas9 is then able to mediate RNA-guided cleavage of a DNA target sequence.

Figure 2 shows the results of T7 Endonuclease I Surveyor assay used to assess ligand-dependent Cas9 gene modification at three target sites (EMX, VEGF, or CLTA). The presence of two bands corresponding to smaller DNA fragments (the fragments are approximately the same size for EMX) indicates genomic modification.

Figure 3A-C. Insertion of an evolved ligand-dependent intein enables small-molecule control of Cas9. (A) Intein insertion renders Cas9 inactive. Upon 4-HT binding, the intein undergoes conformational changes that trigger protein splicing and restore Cas9 activity. (B) The evolved intein was inserted to replace each of the colored residues. Intein-inserted Cas9 variants at S219 and C574 (green) were used in subsequent experiments. (C) Genomic EGFP disruption activity of wild-type Cas9 and intein-Cas9 variants in the absence or presence of 4-HT. Intein-Cas9 variants are identified by the residue replaced by the intein. Error bars reflect the standard deviation of three biological replicates.

Figure 4A-D. Genomic DNA modification by intein-Cas9(S219), intein-Cas9(C574), and wild-type Cas9. (A) Indel frequency from high-throughput DNA sequencing of amplified genomic on-target sites in the absence or presence of 4-HT. Note that a significant number of indels were observed at the CLTA on-target site even in the absence of a targeting sgRNA (Table 9). (B-D) DNA modification specificity, defined as on-target:off-target indel frequency ratio^{45}, normalized to wild-type Cas9. Cells were transfected with 500 ng of the Cas9 expression plasmid. P-values are < 10^{-15} for the Fisher exact test (one-sided up) on comparisons of indel modification frequency in the presence versus the absence of 4-HT for intein-Cas9(S219) and intein-Cas9(C574). P-values were adjusted for multiple comparisons using the Benjamini-Hochberg method, and are listed in Table 5. Error bars reflect the range of two independent experiments conducted on different days.
Figure 5. Effect of 4-HT on cellular toxicity. Untransfected HEK293-GFP stable cells, and cells transfected with intein-Cas9(S219) and sgRNA expression plasmids, were treated with or without 4-HT (1 μM). 12 h after transfection, the media was replaced with full serum media, with or without 4-HT (1 μM). Cells were thus exposed to 4-HT for 0, 12, or 60 h. The live cell population was determined by flow cytometry 60 h after transfection using TO-PRO-3 stain (Life Technologies). Error bars reflect the standard deviation of six technical replicates.

Figure 6A-B. Western blot analysis of HEK293-GFP stable cells transfected with (A) wild-type Cas9 or (B) intein-Cas9(S219) expression plasmid. 12 h after transfection, cells were treated with or without 4-HT (1 μM). Cells were lysed and pooled from three technical replicates 4, 8, 12, or 24 h after 4-HT treatment. An anti-FLAG antibody (Sigma-Aldrich F1804) and an anti-mouse 800CW IRDye (LI-COR) were used to visualize the gel. Lanes 1 and 2 contain purified dCas9-VP64-3xFLAG protein and lysate from untransfected HEK293 cells, respectively.

Figure 7. Indel frequency from high-throughput DNA sequencing of amplified genomic on-target sites ("On") and off-target sites ("Off 1-Off 4") by intein-Cas9(S219), intein-Cas9(C574), and wild-type Cas9 in the presence of 4-HT. 500 ng of Cas9 expression plasmid was transfected. The higher observed efficiency of VEGF Off 1 modification than VEGF on-target modification is consistent with a previous report. P-values are < 0.005 for the Fisher exact test (one-sided down) on all pairwise comparisons within each independent experiment of off-target modification frequency between either intein-Cas9 variant in the presence of 4-HT versus that of wild-type Cas9 in the presence of 4-HT. P-values were adjusted for multiple comparisons using the Benjamini-Hochberg method, and are listed in Table 7. Error bars reflect the range of two independent experiments conducted on different days. See also Fu, Y. et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nature biotechnology 31, 822-826 (2013).

Figure 8A-C. DNA modification specificity of intein-Cas9(S219), intein-Cas9(C574), and wild-type Cas9 in the absence of 4-HT. (A-C) On-target:off-target indel frequency ratio following the transfection of 500 ng of intein-Cas9(S219), intein-Cas9(C574), or wild-type Cas9 expression plasmid.

Figure 9. Genomic on-target DNA modification by intein-Cas9(S219), intein-Cas9(C574), and wild-type Cas9 in the presence of 4-HT. Five different amounts of wild-type Cas9 expression plasmid, specified in parenthesis, were transfected. P-values for
comparisons between conditions (Table 8) were obtained using the Fisher exact test and adjusted for multiple comparisons using the Benjamini-Hochberg Method.

Figure 10A-B. Indel frequency from high-throughput DNA sequencing of amplified genomic on-target sites ("On") and off-target sites ("Off 1-Off 4") by intein-Cas9(S219), intein-Cas9(C574), and wild-type Cas9 in the presence of 4-HT. Five different amounts of wild-type Cas9 expression plasmid, specified in parenthesis, were transfected (A). Genomic sites with low modification frequencies are enlarged in (B). P-values for comparisons between conditions (Table 8) were obtained using the Fisher exact test and adjusted for multiple comparisons using the Benjamini-Hochberg Method.

Figure 11A-C. DNA modification specificity of intein-Cas9(S219), intein-Cas9(C574), and wild-type Cas9 in the presence of 4-HT. (A-C) On-target:off-target indel frequency ratio normalized to wild-type Cas9 (500 ng). Five different amounts of wild-type Cas9 expression plasmid, specified in parenthesis, were transfected.

Figure 12A-B. Genomic EGFP disruption activity of intein-Cas9(S219) and intein-Cas9(S219-G521R) in the presence of (A) β-estradiol or (B) 4-HT. Error bars reflect the standard deviation of three technical replicates.

DEFINITIONS

As used herein and in the claims, the singular forms "a," "an," and "the" include the singular and the plural reference unless the context clearly indicates otherwise. Thus, for example, a reference to "an agent" includes a single agent and a plurality of such agents.

The term "Cas9" or "Cas9 nuclease" refers to an RNA-guided nuclease comprising a Cas9 protein, or a fragment thereof. A Cas9 nuclease is also referred to sometimes as a casnl nuclease or a CRISPR (clustered regularly interspaced short palindromic repeat)-associated nuclease. CRISPR is a prokaryotic adaptive immune system that provides protection against mobile genetic elements (e.g., viruses, transposable elements, and conjugative plasmids). CRISPR clusters contain spacers, sequences complementary to antecedent mobile elements, and target invading nucleic acids. CRISPR clusters are transcribed and processed into CRISPR RNA (crRNA). In type II CRISPR systems correct processing of pre-crRNA requires a trans-encoded small RNA (tracrRNA), endogenous ribonuclease 3 (rnc), and a Cas9 protein. The tracrRNA serves as a guide for ribonuclease 3-aided processing of pre-crRNA. Subsequently, Cas9/crRNA/tracrRNA endonucleolytically cleaves linear or circular dsDNA target complementary to the spacer. The target strand not
complementary to crRNA is first cut endonucleolytically, then trimmed 3′ -> 5′ exonucleolytically. In nature, DNA-binding and cleavage typically requires protein and both RNA species. However, single guide RNAs ("sgRNA", or simply "gRNA") can be engineered so as to incorporate aspects of both the crRNA and tracrRNA into a single RNA molecule. See, e.g., Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. Science 337:816-821(2012), the entire contents of which is hereby incorporated by reference. Cas9 recognizes a short motif in the CRISPR repeat sequences (the PAM or protospacer adjacent motif) to help distinguish self versus non-self. Cas9 nuclease sequences and structures are well known to those of skill in the art (see, e.g., "Complete genome sequence of an M1 strain of Streptococcus pyogenes." Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L. expand/collapse author list McLaughlin R.E., Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma CM., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607(2011); and "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. Science 337:816-821(2012), the entire contents of each of which are incorporated herein by reference). Cas9 orthologs have been described in various species, including, but not limited to, S. pyogenes and S. thermophilus. Additional suitable Cas9 nucleases and sequences will be apparent to those of skill in the art based on this disclosure, and such Cas9 nucleases and sequences include Cas9 sequences from the organisms and loci disclosed in Chylinski, Rhun, and Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) RNA Biology 10:5, 726-737; the entire contents of which are incorporated herein by reference. In some embodiments, proteins comprising Cas9 proteins or fragments thereof are referred to as "Cas9 variants." A Cas9 variant shares homology to Cas9, or a fragment thereof. For example, a Cas9 variant may be at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to wild type Cas9. In some embodiments, the Cas9 variant comprises a fragment of Cas9 (e.g., a gRNA binding domain or a DNA-cleavage domain, an N-terminal domain or a C-terminal domain, etc.), such that the fragment is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5%
identical, or at least about 99.9% to the corresponding fragment of wild type Cas9. In some embodiments, wild type Cas9 corresponds to Cas9 from *Streptococcus pyogenes* (NCBI Reference Sequences: NC_017053.1 and NC_002737.1). In some embodiments, wild type Cas9 corresponds to SEQ ID NO:1 (nucleotide); SEQ ID NO:2 (amino acid)). In some embodiments, Cas9 corresponds to a human codon optimized sequence of Cas9 (e.g., SEQ ID NO:3; See, e.g., Fu et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* 2013; 31, 822-826). In some embodiments, a Cas9 nuclease has an inactive (e.g., an inactivated) DNA cleavage domain. A nuclease-inactivated Cas9 protein may also be referred to as a "dCas9" protein (for nuclease "dead" Cas9). In some embodiments, dCas9 corresponds to, or comprises in part or in whole, the amino acid set forth as SEQ ID NO:5, below. In some embodiments, variants of dCas9 (e.g., variants of SEQ ID NO:5) are provided. For example, in some embodiments, variants having mutations other than D10A and H840A are provided, which e.g., result in a nuclease inactivated Cas9 (dCas9). Such mutations, by way of example, include other amino acid substitutions at D10 and H840, or other substitutions within the nuclease domains of Cas9 (e.g., substitutions in the HNH nuclease subdomain and/or the RuvCl subdomain). In some embodiments, a Cas9 protein variant is a Cas9 nickase, which includes a mutation which abolishes the nuclease activity of one of the two nuclease domains of the protein. In some embodiments, a Cas9 nickase has one, but not both of a D10A and H840A substitution. In some embodiments, a Cas9 nickase corresponds to, or comprises in part or in whole, the amino acid set forth as SEQ ID NO:4, below. In some embodiments, variants or homologues of dCas9 or Cas9 nickase are provided which are at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to SEQ ID NO:5 or SEQ ID NO:4, respectively. In some embodiments, variants of dCas9 or Cas9 nickase (e.g., variants of SEQ ID NO:5 and SEQ ID NO:4, respectively) are provided having amino acid sequences which are shorter, or longer than SEQ ID NO:5 or SEQ ID NO:4, by about 5 amino acids, by about 10 amino acids, by about 15 amino acids, by about 20 amino acids, by about 25 amino acids, by about 30 amino acids, by about 40 amino acids, by about 50 amino acids, by about 75 amino acids, by about 100 amino acids or more.
Cas9; nucleotide (Streptococcus pyogenes)

```
ATGGTACGAGATGAGATGGCAGATTTGGTTATGGTTGACTTTTTAGAAGCTAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAATATACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGAACTTAGCTGGCAGTCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAATTGTTGATGAACTGGTCAAAGTATAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGAATTGCTCAGCTCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGAATCTCATTGCTTTGTCATTGGGATTGACCCCTA
```

(Accession ID: SPP009)
Cas9; amino acid (Streptococcus pyogenes)

MDKK YS IGLDIGTNSVGAWAVITDEYKVPSKKFKVLGNTDHR S IKNKLIGALLFDGDTSEAEATRLKTRA RYTR KRN ICYLIQE IFSFENAKVDSFFHRLEESLVEEDKKHERHP IFGNIDEVAYSHEY KPT YHLRKLVDSTKD R LRLYIYALAMHIFRICHFLIGEDLNPDSVDVFKLFLQVYTNQLFEEFNPASGDVAKAILRSLKSRLEN IAQLPQKEKKNLGNLALSGLTPNFSNFDLAE DAKLQLSKDYDDLNLAAQIQDQADFLAIAKNSLDAI LLSDDLDRVNEITAKASLMISCMIKREVHDHQDLALVRQLEPPKEFIDFSNKGAYIDGASAQEFYFK IKPILEKMDTEELVKNREDLLRKRQTFDNS IPHQILHELHALHLAIRQEDFYPFLKDKNEKIELITFRIP YVGPLARNSRFWMTKRSSEITPWFNEEVVDKGASAQFSFERTMNDKLFNPEVLKHSLLIEYFTFVNYELE KVKVTVEGMRKPFALSEGQIAIVDLFDYQVDFMVRKMAISEEQEIKAHYFSSNNMFKTEI TLANGEIRKPLIEUPGETEIVWDGKRDFAVTKVLSPMQVINKKIBQTVFQFESIELPKRNDSKLARRKDO WPDKKYGFSIPSTVACSMQIKAVSGKKSKSVLKLEGITISSLMERSSKFPQDFLEAQKYEVKVDKLI IKLK IFSELFLENGRRKMLASAGELQKGNELALPVSFNVLYALASHYELKMLSPDEQNLQFLVEQHKHYLDEI IEQI SE FSRKVIWALADNADALYVSNHKKDFKPEIQAGAIENI HILFRTLNLGAPAFAKFDITTDNTRKYSTKEVDATLHIQ SITGLYETRI DLSQ LG (SEQ ID NO:2)

(single underline: HNH domain; double underline: RuvC domain)

dCas9 (DIOA and H840A)

MDKKYSIGAIGTNSVGAWAVITDEYKVPSKKFKVLGNTDHR S IKNKLIGALLFDGDTSEAEATRLKTRA RYTR KRN ICYLIQE IFSFENAKVDSFFHRLEESLVEEDKKHERHP IFGNIDEVAYSHEY KPT YHLRKLVDSTKD R LRLYIYALAMHIFRICHFLIGEDLNPDSVDVFKLFLQVYTNQLFEEFNPASGDVAKAILRSLKSRLEN IAQLPQKEKKNLGNLALSGLTPNFSNFDLAE DAKLQLSKDYDDLNLAAQIQDQADFLAIAKNSLDAI LLSDDLDRVNEITAKASLMISCMIKREVHDHQDLALVRQLEPPKEFIDFSNKGAYIDGASAQEFYFK IKPILEKMDTEELVKNREDLLRKRQTFDNS IPHQILHELHALHLAIRQEDFYPFLKDKNEKIELITFRIP YVGPLARNSRFWMTKRSSEITPWFNEEVVDKGASAQFSFERTMNDKLFNPEVLKHSLLIEYFTFVNYELE KVKVTVEGMRKPFALSEGQIAIVDLFDYQVDFMVRKMAISEEQEIKAHYFSSNNMFKTEI TLANGEIRKPLIEUPGETEIVWDGKRDFAVTKVLSPMQVINKKIBQTVFQFESIELPKRNDSKLARRKDO WPDKKYGFSIPSTVACSMQIKAVSGKKSKSVLKLEGITISSLMERSSKFPQDFLEAQKYEVKVDKLI IKLK IFSELFLENGRRKMLASAGELQKGNELALPVSFNVLYALASHYELKMLSPDEQNLQFLVEQHKHYLDEI IEQI SE FSRKVIWALADNADALYVSNHKKDFKPEIQAGAIENI HILFRTLNLGAPAFAKFDITTDNTRKYSTKEVDATLHIQ SITGLYETRI DLSQ LG (SEQ ID NO:5)

(single underline: HNH domain; double underline: RuvC domain)

Cas9 nickase (Dl OA)(amino acid sequence)

MDKKSYSIGLAIGTNSVGAWAVITDEYKVPSKKFKVLGNTDHR S IKNKLIGALLFDGDTSEAEATRLKTRA RYTR KRN ICYLIQE IFSFENAKVDSFFHRLEESLVEEDKKHERHP IFGNIDEVAYSHEY KPT YHLRKLVDSTKD R LRLYIYALAMHIFRICHFLIGEDLNPDSVDVFKLFLQVYTNQLFEEFNPASGDVAKAILRSLKSRLEN IAQLPQKEKKNLGNLALSGLTPNFSNFDLAE DAKLQLSKDYDDLNLAAQIQDQADFLAIAKNSLDAI LLSDDLDRVNEITAKASLMISCMIKREVHDHQDLALVRQLEPPKEFIDFSNKGAYIDGASAQEFYFK IKPILEKMDTEELVKNREDLLRKRQTFDNS IPHQILHELHALHLAIRQEDFYPFLKDKNEKIELITFRIP YVGPLARNSRFWMTKRSSEITPWFNEEVVDKGASAQFSFERTMNDKLFNPEVLKHSLLIEYFTFVNYELE KVKVTVEGMRKPFALSEGQIAIVDLFDYQVDFMVRKMAISEEQEIKAHYFSSNNMFKTEI TLANGEIRKPLIEUPGETEIVWDGKRDFAVTKVLSPMQVINKKIBQTVFQFESIELPKRNDSKLARRKDO WPDKKYGFSIPSTVACSMQIKAVSGKKSKSVLKLEGITISSLMERSSKFPQDFLEAQKYEVKVDKLI IKLK IFSELFLENGRRKMLASAGELQKGNELALPVSFNVLYALASHYELKMLSPDEQNLQFLVEQHKHYLDEI IEQI SE FSRKVIWALADNADALYVSNHKKDFKPEIQAGAIENI HILFRTLNLGAPAFAKFDITTDNTRKYSTKEVDATLHIQ SITGLYETRI DLSQ LG (SEQ ID NO:5)

(single underline: HNH domain; double underline: RuvC domain)

[0035] Cas9 variants are provided comprising an intein (e.g., a ligand-dependent intein) inserted within the Cas9 sequence and may be referred to as small-molecule-controlled Cas9 or ligand-dependent Cas9. In some embodiments, the intein is inserted into any location (e.g., at any amino acid position) in Cas9. In some embodiments, the inserted intein sequence replaces one or more amino acids in Cas9. For example, in some embodiments the inserted intein sequence replaces any cysteine, any alanine, any threonine, or any serine in Cas9 or a Cas9 variant such as dCas9 or Cas9 nickase. In some embodiments the inserted intein sequence replaces Cys80, Alal27, Thr333, Thr519, Cys574, Thr622, Ser701, Ala728, Thr995, Serl006, Serll54, Serll59, or Serl274 in Cas9 (SEQ ID NO:2), dCas9 (SEQ ID NO:5), or Cas9 nickase (SEQ ID NO:4).

[0036] The terms "conjugating," "conjugated," and "conjugation" refer to an association of two entities, for example, of two molecules such as two proteins, two domains (e.g., a binding domain and a cleavage domain), or a protein and an agent, e.g., a ligand domain and a small molecule. In some aspects, the association is between a protein (e.g., RNA-programmable nuclease) and a nucleic acid (e.g., a guide RNA). The association can be, for example, a direct or indirect (e.g., via a linker) covalent linkage. In some aspects, the association is between two or more proteins, for example, an RNA-programmable nuclease (e.g., Cas9) and an intein protein. In some embodiments, the association is covalent. In some embodiments, two molecules are conjugated via a linker connecting both molecules.

[0037] The term "consensus sequence," as used herein in the context of nucleic acid sequences, refers to a calculated sequence representing the most frequent nucleotide residues found at each position in a plurality of similar sequences. Typically, it represents the results of a multiple sequence alignment in which related sequences are compared to each other and similar sequence motifs are calculated. Methods and software for determining a consensus sequence are known in the art (See, e.g., JalCiew (jalview.org); and UGENE; Okonechnikov, K.; Golosova, O.; Fursov, M.; the UGENE team. "Unipro UGENE: a unified bioinformatics toolkit". Bioinformatics. 2012; doi:10.1093/bioinformatics/bts091).
The term "deaminase" refers to an enzyme that catalyzes a deamination reaction. In some embodiments, the deaminase is a cytidine deaminase, catalyzing the hydrolytic deamination of cytidine or deoxycytidine to uracil or deoxyuracil, respectively.

The term "effective amount," as used herein, refers to an amount of a biologically active agent (e.g., a ligand-dependent Cas9) that is sufficient to elicit a desired biological response. For example, in some embodiments, an effective amount of a nuclease may refer to the amount of the nuclease that is sufficient to induce cleavage of a desired target site-specifically bound and cleaved by the nuclease, preferably with minimal or no off-target cleavage. In some embodiments, an effective amount of another ligand-dependent Cas9 protein having other nucleic acid modifying activities may refer to the amount of the protein that is sufficient to induce the nucleic acid modification. As will be appreciated by the skilled artisan, the effective amount of an agent, e.g., a ligand-dependent nuclease, deaminase, recombinase, nickase, or a hybrid protein, a fusion protein, a protein dimer, a complex of a protein (or protein dimer) and a polynucleotide, or a polynucleotide, may vary depending on various factors as, for example, on the desired biological response, the specific allele, genome, target site, cell, or tissue being targeted, and the agent being used.

The term "engineered," as used herein, refers to a nucleic acid molecule, a protein molecule, complex, substance, or entity that has been designed, produced, prepared, synthesized, and/or manufactured by a human. Accordingly, an engineered product is a product that does not occur in nature.

The term "epigenetic modifier," as used herein, refers to a protein or catalytic domain thereof having enzymatic activity that results in the epigenetic modification of DNA, for example, chromosomal DNA. Epigenetic modifications include, but are not limited to, DNA methylation and demethylation; histone modifications including methylation and demethylation (e.g., mono-, di- and tri-methylation), histone acetylation and deacetylation, as well as histone ubiquitylation, phosphorylation, and sumoylation.

The term "extein," as used herein, refers to a polypeptide sequence that is flanked by an intein and is ligated to another extein during the process of protein splicing to form a mature, spliced protein. Typically, an intein is flanked by two extein sequences that are ligated together when the intein catalyzes its own excision. Exteins, accordingly, are the protein analog to exons found in mRNA. For example, a polypeptide comprising an intein may be of the structure extein(N) - intein - extein(C). After excision of the intein and splicing of the two exteins, the resulting structures are extein(N) - extein(C) and a free intein.
The term "hybrid protein," as used herein, refers to a protein that comprises the amino acid sequence of a target protein (e.g., a Cas9 protein) and, embedded in that amino acid sequence, a ligand-dependent intein as described herein. Accordingly, a hybrid protein generally comprises the structure: target protein(N) - intein - target protein(C). Typically, a hybrid protein comprises a Cas9 protein (e.g., Cas9, Cas9 variants such as dCas9, fragments of Cas9 or Cas9 variants, etc.) and a ligand-dependent intein. In some embodiments, a hybrid protein is encoded by a recombinant nucleic acid, in which a nucleic acid sequence encoding an intein is inserted in frame into a nucleic acid sequence encoding a target protein. In certain embodiments, the target protein exhibits a desired activity or property that is absent or reduced in the hybrid protein. In some embodiments, excision of the intein from the hybrid protein results in a restoration of the desired activity or property in the mature, spliced target protein. Non-limiting examples of desired activities or properties of target proteins are binding activities, enzymatic activities (e.g., nuclease activities, gene editing activities, deaminase activities, recombinase activities), reporter activities (e.g., fluorescent activity), therapeutic activities, size, charge, hydrophobicity, hydrophilicity, or 3D-structure. In some embodiments, excision of the intein from a hybrid protein results in a mature, spliced target protein that exhibits the same or similar levels of a desired activity as the native target protein. A hybrid protein may be created from any target protein by embedding an intein sequence into the amino acid sequence of the target protein, for example, by generating a recombinant, hybrid protein-encoding nucleic acid molecule and subsequent transcription and translation, or by protein synthesis methods known to those of skill in the art.

The term "intein," as used herein, refers to an amino acid sequence that is able to excise itself from a protein and to rejoin the remaining protein segments (the exteins) via a peptide bond in a process termed protein splicing. Inteins are analogous to the introns found in mRNA. Many naturally occurring and engineered inteins and hybrid proteins comprising such inteins are known to those of skill in the art, and the mechanism of protein splicing has been the subject of extensive research. As a result, methods for the generation of hybrid proteins from naturally occurring and engineered inteins are well known to the skilled artisan. For an overview, see pages 1-10, 193-207, 211-229, 233-252, and 325-341 of Gross, Belfort, Derbyshire, Stoddard, and Wood (Eds.) Homing Endonucleases and Inteins Springer Verlag Heidelberg, 2005; ISBN 9783540251064; the contents of which are incorporated herein by reference for disclosure of inteins and methods of generating hybrid proteins comprising natural or engineered inteins. As will be apparent to those of skill in the art, an intein may
catalyze protein splicing in a variety of extein contexts. Accordingly, an intein can be introduced into virtually any target protein sequence to create a desired hybrid protein, and the invention is not limited in the choice of target proteins.

The term "intein domain," as used herein, refers to the amino acid sequence of an intein that is essential for self-excision and extein ligation. For example, in some inteins, the entire intein amino acid sequence, or part(s) thereof, may constitute the intein domain, while in ligand-dependent inteins, the ligand-binding domain is typically embedded into the intein domain, resulting in the structure: intein domain (N) - ligand-binding domain - intein domain (C).

The term "ligand binding domain," as used herein, refers to a peptide or protein domain that binds a ligand. A ligand binding domain may be a naturally occurring domain or an engineered domain. Examples of ligand-binding domains referred to herein are the ligand binding domain of a native estrogen receptor, e.g., the ligand-binding domain of the native human estrogen receptor, and engineered, evolved, or mutated derivatives thereof. Other suitable ligand binding domains include the human thyroid hormone receptor (see, e.g., Skretas et al., "Regulation of protein activity with small-molecule-controlled inteins." Protein Sci. 2005; 14, 523-532) and members of the ribose-binding protein family (see, e.g., Bjorkman et al., "Multiple open forms of ribose-binding protein trace the path of its conformational change." J Mol Biol. 1998 12;279(3):651-64). Typically, a ligand-binding domain useful in the context of ligand-dependent inteins, as provided herein, exhibits a specific three-dimensional structure in the absence of the ligand, which inhibits intein self-excision, and undergoes a conformational change upon binding of the ligand, which promotes intein self-excision. Some of the ligand-dependent inteins provided herein comprise a ligand-binding domain derived from the estrogen receptor that can bind 4-HT and other estrogen-receptor ligands, e.g., ligands described in more detail elsewhere herein, and undergo a conformational change upon binding of the ligand. An appropriate ligand may be any chemical compound that binds the ligand-binding domain and induces a desired conformational change. In some embodiments, an appropriate ligand is a molecule that is bound by the ligand-binding domain with high specificity and affinity. In some embodiments, the ligand is a small molecule. In some embodiments, the ligand is a molecule that does not naturally occur in the context (e.g., in a cell or tissue) that a ligand-dependent intein is used in. For example, in some embodiments, the ligand-binding domain is a ligand-binding domain derived from an estrogen receptor, and the ligand is tamoxifen, or a derivative or analog thereof (e.g., 4-hydroxytamoxifen, 4-HT).
The term "ligand-dependent intein," as used herein refers to an intein that comprises a ligand-binding domain. Typically, the ligand-binding domain is inserted into the amino acid sequence of the intein, resulting in a structure intein (N) - ligand-binding domain - intein (C). Typically, ligand-dependent inteins exhibit no or only minimal protein splicing activity in the absence of an appropriate ligand, and a marked increase of protein splicing activity in the presence of the ligand. In some embodiments, the ligand-dependent intein does not exhibit observable splicing activity in the absence of ligand but does exhibit splicing activity in the presence of the ligand. In some embodiments, the ligand-dependent intein exhibits an observable protein splicing activity in the absence of the ligand, and a protein splicing activity in the presence of an appropriate ligand that is at least 5 times, at least 10 times, at least 50 times, at least 100 times, at least 150 times, at least 200 times, at least 250 times, at least 500 times, at least 1000 times, at least 1500 times, at least 2000 times, at least 2500 times, at least 5000 times, at least 10000 times, at least 20000 times, at least 25000 times, at least 50000 times, at least 100000 times, or at least 1000000 times greater than the activity observed in the absence of the ligand. In some embodiments, the increase in activity is dose dependent over at least 1 order of magnitude, at least 2 orders of magnitude, at least 3 orders of magnitude, at least 4 orders of magnitude, or at least 5 orders of magnitude, allowing for fine-tuning of intein activity by adjusting the concentration of the ligand. Suitable ligand-dependent inteins are known in the art, and in include those provided below and those described in published U.S. Patent Application U.S. 2014/0065711 Al; Mootz et al., "Protein splicing triggered by a small molecule." J. Am. Chem. Soc. 2002; 124, 9044-9045; Mootz et al., "Conditional protein splicing: a new tool to control protein structure and function in vitro and in vivo." J. Am. Chem. Soc. 2003; 125, 10561-10569; Buskirk et al., Proc. Natl. Acad. Sci. USA. 2004; 101, 10505-10510); Skretas & Wood, "Regulation of protein activity with small-molecule-controlled inteins." Protein Sci. 2005; 14, 523-532; Schwartz, et al., "Post-translational enzyme activation in an animal via optimized conditional protein splicing." Nat. Chem. Biol. 2007; 3, 50-54; Peck et al., Chem. Biol. 2011; 18 (5), 619-630; the entire contents of each are hereby incorporated by reference.

2-4 intein:
CLAEGTRIFDPVTGTTHIEDVVDGRKP HJVVAAKGDGLLARPVVSFWFDQGTRDVIGLRIAGGAIJWATPDHKYLITEYGWRAAGELKRDLRVAEGGGSNSLALSHTAQMVSALLDAEPPLYSEYDTPSFEASMMGLLTNLADRELVHMINKWAKRVPFGVDLTLHDQAHLECAWLELMGILVWRSMEHGKLFFAPNLLLDNRQG KCVEGMVE IFDML LATS SRFMNMILQEGEEFCVCLS 11 LLSNGVYTFLS STLKSLIEKDHIRALDKI TDTL IHLMKAQGLTLQQHQR LAQLLL 1LSHIRHMSNKGEHLYSMKYNVPLYDLLEMLDAHRLHAGGGASRVQAFADALDDKFLHDMLAEE LRYSVIREVLPRARRTFDLEVEELHTLVAEGVVVHNC (SEQ ID NO:7)
3-2 intein:
CLAEGRTRIFDPVTGTTHIREDVVDGRKPIHVVAADGTLARRVPSWFDQGTRDVGRIAGGAIVWATPDHKV
LTEYGWRAAGELRKGDRVAGPGGGSNLALSITADQVMASALLDEPPILYSEYDTPSFSEAMMGLLTNLADRE
LVHMNVNARKVPGFVDTLHDQAHLLECAWLEIIMGLVWRMSMEHPGKLLFAPNLLDLRNQGKCVEGMVEIFDML
LATSSRFMMNLQGEEFVCLKS ILLNSGVYTFSLSTLKSLEKDHHRALDKITDLTILHMAKLTLQQHQQR
LQAQLLIIILHRIHMSNKMEHLYSMKYNVPLYLDDLLEMLDARHLHGSGASRVQAFADALDDKFKLFHDLMAE
LRYSVIREVLPTRRARTDFDEVEELHTLVAEGVVVHCN (SEQ ID NO:8)

30R3-1 intein:
CLAEGRTRIFDPVTGTTHIREDVVDGRKPIHVVAADGTLARRVPSWFDQGTRDVGRIAGGAIVWATPDHKV
LTEYGWRAAGELRKGDRVAGPGGGSNLALSITADQVMASALLDEPPILYSEYDTPSFSEAMMGLLTNLADRE
LVHMNVNARKVPGFVDTLHDQAHLLECAWLEIIMGLVWRMSMEHPGKLLFAPNLLDLRNQGKCVEGMVEIFDML
LATSSRFMMNLQGEEFVCLKS ILLNSGVYTFSLSTLKSLEKDHHRALDKITDLTILHMAKLTLQQHQQR
LQAQLLIIILHRIHMSNKMEHLYSMKYNVPLYLDDLLEMLDARHLHGSGASRVQAFADALDDKFKLFHDLMAE
LRYSVIREVLPTRRARTDFDEVEELHTLVAEGVVVHCN (SEQ ID NO:9)

30R3-2 intein:
CLAEGRTRIFDPVTGTTHIREDVVDGRKPIHVVAADGTLARRVPSWFDQGTRDVGRIAGGAIVWATPDHKV
LTEYGWRAAGELRKGDRVAGPGGGSNLALSITADQVMASALLDEPPILYSEYDTPSFSEAMMGLLTNLADRE
LVHMNVNARKVPGFVDTLHDQAHLLECAWLEIIMGLVWRMSMEHPGKLLFAPNLLDLRNQGKCVEGMVEIFDML
LATSSRFMMNLQGEEFVCLKS ILLNSGVYTFSLSTLKSLEKDHHRALDKITDLTILHMAKLTLQQHQQR
LQAQLLIIILHRIHMSNKMEHLYSMKYNVPLYLDDLLEMLDARHLHGSGASRVQAFADALDDKFKLFHDLMAE
LRYSVIREVLPTRRARTDFDEVEELHTLVAEGVVVHCN (SEQ ID NO:10)

30R3-3 intein:
CLAEGRTRIFDPVTGTTHIREDVVDGRKPIHVVAADGTLARRVPSWFDQGTRDVGRIAGGAIVWATPDHKV
LTEYGWRAAGELRKGDRVAGPGGGSNLALSITADQVMASALLDEPPILYSEYDTPSFSEAMMGLLTNLADRE
LVHMNVNARKVPGFVDTLHDQAHLLECAWLEIIMGLVWRMSMEHPGKLLFAPNLLDLRNQGKCVEGMVEIFDML
LATSSRFMMNLQGEEFVCLKS ILLNSGVYTFSLSTLKSLEKDHHRALDKITDLTILHMAKLTLQQHQQR
LQAQLLIIILHRIHMSNKMEHLYSMKYNVPLYLDDLLEMLDARHLHGSGASRVQAFADALDDKFKLFHDLMAE
LRYSVIREVLPTRRARTDFDEVEELHTLVAEGVVVHCN (SEQ ID NO:11)

37R3-1 intein:
CLAEGRTRIFDPVTGTTHIREDVVDGRKPIHVVAADGTLARRVPSWFDQGTRDVGRIAGGAIVWATPDHKV
LTEYGWRAAGELRKGDRVAGPGGGSNLALSITADQVMASALLDEPPILYSEYDTPSFSEAMMGLLTNLADRE
LVHMNVNARKVPGFVDTLHDQAHLLECAWLEIIMGLVWRMSMEHPGKLLFAPNLLDLRNQGKCVEGMVEIFDML
LATSSRFMMNLQGEEFVCLKS ILLNSGVYTFSLSTLKSLEKDHHRALDKITDLTILHMAKLTLQQHQQR
LQAQLLIIILHRIHMSNKMEHLYSMKYNVPLYLDDLLEMLDARHLHGSGASRVQAFADALDDKFKLFHDLMAE
LRYSVIREVLPTRRARTDFDEVEELHTLVAEGVVVHCN (SEQ ID NO:12)

37R3-2 intein:
CLAEGRTRIFDPVTGTTHIREDVVDGRKPIHVVAADGTLARRVPSWFDQGTRDVGRIAGGAIVWATPDHKV
LTEYGWRAAGELRKGDRVAGPGGGSNLALSITADQVMASALLDEPPILYSEYDTPSFSEAMMGLLTNLADRE
LVHMNVNARKVPGFVDTLHDQAHLLECAWLEIIMGLVWRMSMEHPGKLLFAPNLLDLRNQGKCVEGMVEIFDML
LATSSRFMMNLQGEEFVCLKS ILLNSGVYTFSLSTLKSLEKDHHRALDKITDLTILHMAKLTLQQHQQR
LQAQLLIIILHRIHMSNKMEHLYSMKYNVPLYLDDLLEMLDARHLHGSGASRVQAFADALDDKFKLFHDLMAE
LRYSVIREVLPTRRARTDFDEVEELHTLVAEGVVVHCN (SEQ ID NO:13)
The term "linker," as used herein, refers to a chemical group or a molecule linking two adjacent molecules or moieties, *e.g.*, two polypeptides. Typically, the linker is positioned between, or flanked by, two groups, molecules, or other moieties and connected to each one via a covalent bond, thus connecting the two. In some embodiments, the linker is an amino acid linker. In some embodiments, the amino acid linker comprises at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, or at least 30 amino acids. In some embodiments, the linker is a divalent organic molecule, group, polymer, or chemical moiety. In some embodiments, the peptide linker comprises repeats of the tri-peptide Gly-Gly-Ser, *e.g.*, comprising the sequence (GGS)$_n$, wherein $n$ represents at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more repeats. In some embodiments, the linker comprises the sequence (GGS)$_6$ (SEQ ID NO: 15). In some embodiments, the peptide linker is the 16 residue "XTEN" linker, or a variant thereof (See, *e.g.*, Schellenberger *et al.* A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. *Nat. Biotechnol.* 27, 1186-1 190 (2009)). In some embodiments, the XTEN linker comprises the sequence SGSETPGTSESATPES (SEQ ID NO: 16), SGSETPGTSESA (SEQ ID NO: 17), or SGSETPGTSESATPEGGS (SEQ ID NO: 18). In some embodiments, the peptide linker is one or more selected from VPFLLEPDNINGKTC (SEQ ID NO: 19), GSAGSAAGSGEF (SEQ ID NO:20), SIVAQLSRPDPA (SEQ ID NO:21), MKIIEQLPSA (SEQ ID NO:22), VRHKLKRVG (SEQ ID NO:23), GHGTGSTGSGSS (SEQ ID NO:24), MSRDPDA (SEQ ID NO:25); or GGGM (SEQ ID NO:26).

The term "mutation," as used herein, refers to a substitution of a residue within a sequence, *e.g.*, a nucleic acid or amino acid sequence, with another residue, or a deletion or insertion of one or more residues within a sequence. Mutations are typically described herein by identifying the original residue followed by the position of the residue within the sequence and by the identity of the newly substituted residue. Methods for making the amino acid substitutions (mutations) provided herein are known in the art and are provided by, for example, Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)).

The term "nuclease," as used herein, refers to an agent, for example, a protein, capable of cleaving a phosphodiester bond connecting two nucleotide residues in a nucleic acid.
acid molecule. In some embodiments, "nuclease" refers to a protein having an inactive DNA cleavage domain, such that the nuclease is incapable of cleaving a phosphodiester bond. In some embodiments, the nuclease is a protein, e.g., an enzyme that can bind a nucleic acid molecule and cleave a phosphodiester bond connecting nucleotide residues within the nucleic acid molecule. A nuclease may be an endonuclease, cleaving a phosphodiester bonds within a polynucleotide chain, or an exonuclease, cleaving a phosphodiester bond at the end of the polynucleotide chain. In some embodiments, a nuclease is a site-specific nuclease, binding and/or cleaving a specific phosphodiester bond within a specific nucleotide sequence, which is also referred to herein as the "recognition sequence," the "nuclease target site," or the "target site." In some embodiments, a nuclease is an RNA-guided (i.e., RNA-programmable) nuclease, which is associated with (e.g., binds to) an RNA (e.g., a guide RNA, "gRNA") having a sequence that complements a target site, thereby providing the sequence specificity of the nuclease. In some embodiments, a nuclease recognizes a single stranded target site. In some embodiments, a nuclease recognizes a double-stranded target site, for example, a double-stranded DNA target site. The target sites of many naturally occurring nucleases, for example, many naturally occurring DNA restriction nucleases, are well known to those of skill in the art. In many cases, a DNA nuclease, such as EcoRI, HindIII, or BamHI, recognize a palindromic, double-stranded DNA target site of 4 to 10 base pairs in length, and cut each of the two DNA strands at a specific position within the target site. Some endonucleases cut a double-stranded nucleic acid target site symmetrically, i.e., cutting both strands at the same position so that the ends comprise base-paired nucleotides, also referred to herein as blunt ends. Other endonucleases cut a double-stranded nucleic acid target site asymmetrically, i.e., cutting each strand at a different position so that the ends include unpaired nucleotides. Unpaired nucleotides at the end of a double-stranded DNA molecule are also referred to as "overhangs," e.g., as "5'-overhang" or as "3'-overhang," depending on whether the unpaired nucleotide(s) form(s) the 5' or the 3' end of the respective DNA strand. Double-stranded DNA molecule ends ending with unpaired nucleotide(s) are also referred to as sticky ends, as they can "stick to" other double-stranded DNA molecule ends comprising complementary unpaired nucleotide(s). A nuclease protein typically comprises a "binding domain" that mediates the interaction of the protein with the nucleic acid substrate, and also, in some cases, specifically binds to a target site, and a "cleavage domain" that catalyzes the cleavage of the phosphodiester bond within the nucleic acid backbone. In some embodiments, a nuclease protein can bind and cleave a nucleic acid molecule in a monomeric form. In some embodiments, a nuclease protein has to dimerize or multimerize in order to
cleave a target nucleic acid. Binding domains and cleavage domains of naturally occurring nuclease, as well as modular binding domains and cleavage domains that can be fused to create nuclease binding specific target sites, are well known to those of skill in the art. For example, the binding domain of RNA-programmable nuclease (e.g., Cas9), or a Cas9 protein having an inactive DNA cleavage domain (e.g., dCas9), can be used as a binding domain (e.g., that binds a gRNA to direct binding to a target site) to specifically bind a desired target site, and fused or conjugated to a cleavage domain, for example, the cleavage domain of FokI, to create an engineered nuclease cleaving the target site. In some embodiments, Cas9 fusion proteins provided herein comprise the cleavage domain of FokI, and are therefore referred to as "fCas9" proteins. In some embodiments, the cleavage domain of FokI, e.g., in a fCas9 protein corresponds to, or comprises in part or whole, the amino acid sequence (or variants thereof) set forth as SEQ ID NO:6, below. In some embodiments, variants or homologues of the FokI cleavage domain include any variant or homologue capable of dimerizing (e.g., as part of fCas9 fusion protein) with another FokI cleavage domain at a target site in a target nucleic acid, thereby resulting in cleavage of the target nucleic acid. In some embodiments, variants of the FokI cleavage domain (e.g., variants of SEQ ID NO:6) are provided which are at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to SEQ ID NO:6. In some embodiments, variants of the FokI cleavage domain (e.g., variants of SEQ ID NO:6) are provided having an amino acid sequence which is shorter, or longer than SEQ ID NO:6, by about 5 amino acids, by about 10 amino acids, by about 15 amino acids, by about 20 amino acids, by about 25 amino acids, by about 30 amino acids, by about 40 amino acids, by about 50 amino acids, by about 75 amino acids, by about 100 amino acids, or more.

Cleavage domain of FokI:

GSQVKSEEEKKSELRLKLYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLLGGSRKPGAIYTV
GSPIDYGVIKAYSGGYNLPIGQADEMQRYVEENQTRNHIPNEWKPVPSVTEFKPLVSGHFKGNYKAQ
LTRNLHITNCNAGVLSVEELIGEMIKAGTLTLEEVRRKFNNGEINF (SEQ ID NO:6)

[0051] The terms "nucleic acid" and "nucleic acid molecule," as used herein, refer to a compound comprising a nucleobase and an acidic moiety, e.g., a nucleoside, a nucleotide, or a polymer of nucleotides. Typically, polymeric nucleic acids, e.g., nucleic acid molecules comprising three or more nucleotides are linear molecules, in which adjacent nucleotides are linked to each other via a phosphodiester linkage. In some embodiments, the term "nucleic
"acid" refers to individual nucleic acid residues (e.g. nucleotides and/or nucleosides). In some embodiments, "nucleic acid" refers to an oligonucleotide chain comprising three or more individual nucleotide residues. As used herein, the terms "oligonucleotide" and "polynucleotide" can be used interchangeably to refer to a polymer of nucleotides (e.g., a string of at least three nucleotides). In some embodiments, "nucleic acid" encompasses RNA as well as single and/or double-stranded DNA. Nucleic acids may be naturally occurring, for example, in the context of a genome, a transcript, an mRNA, tRNA, rRNA, siRNA, snRNA, a plasmid, cosmid, chromosome, chromatid, or other naturally occurring nucleic acid molecule. On the other hand, a nucleic acid molecule may be a non-naturally occurring molecule, e.g., a recombinant DNA or RNA, an artificial chromosome, an engineered genome, or fragment thereof, or a synthetic DNA, RNA, DNA/RNA hybrid, or including non-naturally occurring nucleotides or nucleosides. Furthermore, the terms "nucleic acid," "DNA," "RNA," and/or similar terms include nucleic acid analogs, i.e. analogs having other than a phosphodiester backbone. Nucleic acids can be purified from natural sources, produced using recombinant expression systems and optionally purified, chemically synthesized, etc. Where appropriate, e.g., in the case of chemically synthesized molecules, nucleic acids can comprise nucleoside analogs such as analogs having chemically modified bases or sugars, and backbone modifications. A nucleic acid sequence is presented in the 5' to 3' direction unless otherwise indicated. In some embodiments, a nucleic acid is or comprises natural nucleosides (e.g. adenosine, thymidine, guanosine, cytidine, uridine, deoxyadenosine, deoxythymidine, deoxyguanosine, and deoxycytidine); nucleoside analogs (e.g., 2-aminoadenosine, 2-thiothymidine, inosine, pyrrolo-pyrimidine, 3-methyl adenosine, 5-methylcytidine, 2-aminoadenosine, C5-bromouridine, C5-fluorouridine, C5-iodouridine, C5-propynyl-uridine, C5-propynyl-cytidine, C5-methylcytidine, 2-aminoadenosine, 7-deazaadenosine, 7-deazaguanosine, 8-oxoadenosine, 8-oxoguanosine, O(6)-methylguanine, and 2-thiocytidine); chemically modified bases; biologically modified bases (e.g., methylated bases); intercalated bases; modified sugars (e.g., 2'-fluororibose, ribose, 2'-deoxyribose, arabinose, and hexose); and/or modified phosphate groups (e.g., phosphorothioates and 5'-N-phosphoramidite linkages).

The term "pharmaceutical composition," as used herein, refers to a composition that can be administrated to a subject in the context of treatment of a disease or disorder. In some embodiments, a pharmaceutical composition comprises an active ingredient, e.g., a nuclease or a nucleic acid encoding a nuclease, and a pharmaceutically acceptable excipient.
The term "proliferative disease," as used herein, refers to any disease in which cell or tissue homeostasis is disturbed in that a cell or cell population exhibits an abnormally elevated proliferation rate. Proliferative diseases include hyperproliferative diseases, such as pre-neoplastic hyperplastic conditions and neoplastic diseases. Neoplastic diseases are characterized by an abnormal proliferation of cells and include both benign and malignant neoplasias. Malignant neoplasia is also referred to as cancer.

The terms "protein," "peptide," and "polypeptide" are used interchangeably herein and refer to a polymer of amino acid residues linked together by peptide (amide) bonds. The terms refer to a protein, peptide, or polypeptide of any size, structure, or function. Typically, a protein, peptide, or polypeptide will be at least three amino acids long. A protein, peptide, or polypeptide may refer to an individual protein or a collection of proteins. One or more of the amino acids in a protein, peptide, or polypeptide may be modified, for example, by the addition of a chemical entity such as a carbohydrate group, a hydroxyl group, a phosphate group, a farnesyl group, an isofarnesyl group, a fatty acid group, a linker for conjugation, functionalization, or other modification, etc. A protein, peptide, or polypeptide may also be a single molecule or may be a multi-molecular complex. A protein, peptide, or polypeptide may be just a fragment of a naturally occurring protein or peptide. A protein, peptide, or polypeptide may be naturally occurring, recombinant, or synthetic, or any combination thereof.

The term "fusion protein" as used herein refers to a hybrid polypeptide which comprises protein domains from at least two different proteins. One protein may be located at the amino-terminal (N-terminal) portion of the fusion protein or at the carboxy-terminal (C-terminal) protein thus forming an "amino-terminal fusion protein" or a "carboxy-terminal fusion protein," respectively. A protein may comprise different domains, for example, a nucleic acid binding domain (e.g., the gRNA binding domain of Cas9 that directs the binding of the protein to a target site or a dCas9 protein) and a nucleic acid cleavage domain(s). In some embodiments, a protein is in a complex with, or is in association with, a nucleic acid, e.g., DNA or RNA. Any of the proteins provided herein may be produced by any method known in the art. For example, the proteins provided herein may be produced via recombinant protein expression and purification, which is especially suited for fusion proteins comprising a peptide linker. Methods for recombinant protein expression and purification are well known and include those described by Green and Sambrook, Molecular Cloning: A Laboratory Manual (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)), the entire contents of which are incorporated herein by reference.
[0056] The term "protein splicing," as used herein, refers to a process in which a sequence, an intein, is excised from within an amino acid sequence, and the remaining fragments of the amino acid sequence, the exteins, are ligated via an amide bond to form a continuous amino acid sequence.

[0057] The term "RNA-programmable nuclease" and "RNA-guided nuclease" are used interchangeably herein and refer to a nuclease that forms a complex with (e.g., binds or associates with) one or more RNAs that is not a target for cleavage. In some embodiments, an RNA-programmable nuclease, when in a complex with an RNA, may be referred to as a nuclease:RNA complex. Typically, the bound RNA(s) is referred to as a guide RNA (gRNA). gRNAs can exist as an association of two or more RNAs, or as a single RNA molecule. gRNAs that exist as a single RNA molecule may be referred to as single-guide RNAs (sgRNAs), though "gRNA" is used interchangeably to refer to guide RNAs that exist as either a single molecule or as a complex of two or more molecules. Typically, gRNAs that exist as single RNA species comprise at least two domains: (1) a domain that shares homology to a target nucleic acid and may direct binding of a Cas9 complex to the target; and (2) a domain that binds a Cas9 protein. In some embodiments, domain (2) corresponds to a sequence known as a tracrRNA and comprises a stem-loop structure. For example, in some embodiments, domain (2) is homologous to a tracrRNA as depicted in Figure IE of Jinek et al, Science 337:816-821(2012), the entire contents of which is incorporated herein by reference. In some embodiments, domain 2 is at least 90%, at least 95%, at least 98%, or at least 99% identical to the tracrRNA as described by Jinek et al, Science 337:816-821(2012). The gRNA comprises a nucleotide sequence that complements a target site, which mediates binding of the nuclease/RNA complex to said target site and the sequence specificity of the nuclease:RNA complex. The sequence of a gRNA that binds a target nucleic acid can comprise any sequence that complements a region of the target and is suitable for a nuclease:RNA complex to bind. In some embodiments, the RNA-programmable nuclease is the (CRISPR-associated system) Cas9 endonuclease, for example, Cas9 (Csn1) from Streptococcus pyogenes (see, e.g., "Complete genome sequence of an M1 strain of Streptococcus pyogenes." Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L. expand/collapse author list McLaughlin R.E., Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma CM., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607(2011);

Because RNA-programmable nucleases (e.g., Cas9) use RNA:DNA hybridization to determine target DNA cleavage sites, these proteins are able to cleave, in principle, any sequence specified by the guide RNA. Methods of using RNA-programmable nucleases, such as Cas9, for site-specific cleavage (e.g., to modify a genome) are known in the art (see e.g., Cong et al. Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819-823 (2013); Mali et al. RNA-guided human genome engineering via Cas9. Science 339, 823-826 (2013); Hwang et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nature biotechnology 31, 227-229 (2013); Jinek et al. RNA-programmed genome editing in human cells. eLife 2, e00471 (2013); Dicarlo et al. Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic acids research (2013); Jiang et al. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nature biotechnology 31, 233-239 (2013); the entire contents of each of which are incorporated herein by reference).

The term "recombinase," as used herein, refers to a site-specific enzyme that mediates the recombination of DNA between recombinase recognition sequences, which results in the excision, integration, inversion, or exchange (e.g., translocation) of DNA fragments between the recombinase recognition sequences. Recombinases can be classified into two distinct families: serine recombinases (e.g., resolvases and invertases) and tyrosine recombinases (e.g., integrases). Examples of serine recombinases include, without limitation, Hin, Gin, Tn3, β-six, CinH, ParA, γ8, Bxb1, 4>C31, TP901, TGI, φBTI, R4, cpRVI, cpFCl, MRU, A118, U153, and gp29. Examples of tyrosine recombinases include, without limitation, Cre, FLP, R, Lambda, HK101, HK022, and pSAM2. The serine and tyrosine recombinase names stem from the conserved nucleophilic amino acid residue that the recombinase uses to attack the DNA and which becomes covalently linked to the DNA during strand exchange. Recombinases have numerous applications, including the creation of gene knockouts/knock-ins and gene therapy applications. See, e.g., Brown et al., "Serine recombinases as tools for genome engineering." Methods. 201 1:53(4):372-9; Hirano et al., "Site-specific recombinases as tools for heterologous gene integration." Appl. Microbiol. Biotechnol. 2011; 92(2):227-39; Chavez and Calos, "Therapeutic applications of the OC31 integrase system." Curr. Gene Ther. 2011;11(5):375-81; Turan and Bode, "Site-specific
recombinases: from tag-and-target- to tag-and-exchange-based genomic modifications.”
FASEB J. 2011; 25(12):4088-107; Venken and Bellen, "Genome-wide manipulations of Drosophila melanogaster with transposons, Flp recombinase, and OC31 integrase." Methods Mol. Biol. 2012; 859:203-28; Murphy, "Phage recombinases and their applications." Adv. Virus Res. 2012; 83:367-414; Zhang et al., "Conditional gene manipulation: Creating a new biological era." J. Zhejiang Univ. Sci. B. 2012; 13(7):51 1-24; Karpenshif and Bernstein, "From yeast to mammals: recent advances in genetic control of homologous recombination." DNA Repair (Amst). 2012; 11(10):781-8; the entire contents of each are hereby incorporated by reference in their entirety. The recombinases provided herein are not meant to be exclusive examples of recombinases that can be used in embodiments of the invention. The methods and compositions of the invention can be expanded by mining databases for new orthogonal recombinases or designing synthetic recombinases with defined DNA specificities (See, e.g., Groth et al., "Phage integrases: biology and applications." J. Mol. Biol. 2004; 335, 667-678; Gordley et al., "Synthesis of programmable integrases." Proc. Natl. Acad. Sci. USA. 2009; 106, 5053-5058; the entire contents of each are hereby incorporated by reference in their entirety). Other examples of recombinases that are useful in the methods and compositions described herein are known to those of skill in the art, and any new recombinase that is discovered or generated is expected to be able to be used in the different embodiments of the invention. In some embodiments, the catalytic domains of a recombinase are fused to a nuclease-inactivated RNA-programmable nuclease (e.g., dCas9, or a fragment thereof), such that the recombinase domain does not comprise a nucleic acid binding domain or is unable to bind to a target nucleic acid (e.g., the recombinase domain is engineered such that it does not have specific DNA binding activity). Recombinases lacking DNA binding activity and methods for engineering such are known, and include those described by Klippel et al., "Isolation and characterisation of unusual gin mutants." EMBO J. 1988; 7: 3983-3989: Burke et al., "Activating mutations of Tn3 resolvase marking interfaces important in recombination catalysis and its regulation. Mol Microbiol. 2004; 51: 937-948; Olorunni et al., "Synapsis and catalysis by activated Tn3 resolvase mutants." Nucleic Acids Res. 2008; 36: 7181-7191; Rowland et al., "Regulatory mutations in Sin recombinase support a structure-based model of the synaptosome." Mol Microbiol. 2009; 74: 282-298; Akopian et al., "Chimeric recombinases with designed DNA sequence recognition." Proc Natl Acad Sci USA. 2003;100: 8688-8691; Gordley et al., "Evolution of programmable zinc finger-recombinases with activity in human cells. J Mol Biol. 2007; 367: 802-813; Gordley et al., "Synthesis of programmable integrases." Proc Natl Acad Sci USA. 2009; 106: 5053-
5058; Arnold et al, “Mutants of TnJ resolvase which do not require accessory binding sites for recombination activity.” EMBO J. 1999;18: 1407-1414; Gaj et al, "Structure-guided reprogramming of serine recombinase DNA sequence specificity." Proc Natl Acad Sci USA. 2011;108(2):498-503; and Proudfoot et al, "Zinc finger recombinases with adaptable DNA sequence specificity." PLoS One. 2011;6(4):e19537; the entire contents of each are hereby incorporated by reference. For example, serine recombinases of the resolvase-invertase group, e.g., Tn3 and γδ resolvases and the Hin and Gin invertases, have modular structures with autonomous catalytic and DNA-binding domains (See, e.g., Grindley et al, "Mechanism of site-specific recombination." Ann Rev Biochem. 2006; 75: 567-605, the entire contents of which are incorporated by reference). The catalytic domains of these recombinases are thus amenable to being recombined with nuclease-inactivated RNA-programmable nucleases (e.g., dCas9, or a fragment thereof) as described herein, e.g., following the isolation of 'activated' recombinase mutants which do not require any accessory factors (e.g., DNA binding activities) (See, e.g., Klippel et al, "Isolation and characterisation of unusual gin mutants." EMBO J. 1988; 7; 3983-3989; Burke et al, "Activating mutations of Tn3 resolvase marking interfaces important in recombination catalysis and its regulation. Mol Microbiol. 2004; 51: 937-948; Olorunniji et al, "Synapsis and catalysis by activated Tn3 resolvase mutants." Nucleic Acids Res. 2008; 36: 7181-7191; Rowland et al, "Regulatory mutations in Sin recombinase support a structure-based model of the synaptosome." Mol Microbiol. 2009; 74: 282-298; Akopian et al, "Chimeric recombinases with designed DNA sequence recognition." Proc Natl Acad Sci USA. 2003;100: 8688-8691). Additionally, many other natural serine recombinases having an N-terminal catalytic domain and a C-terminal DNA binding domain are known (e.g., phiC31 integrase, TnpX transposase, IS607 transposase), and their catalytic domains can be co-opted to engineer programmable site-specific recombinases as described herein (See, e.g., Smith et al, "Diversity in the serine recombinases." Mol Microbiol. 2002;44: 299-307, the entire contents of which are incorporated by reference). Similarly, the core catalytic domains of tyrosine recombinases (e.g., Cre, λ integrase) are known, and can be similarly co-opted to engineer programmable site-specific recombinases as described herein (See, e.g., Guo et al, "Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse." Nature. 1997; 389:40-46; Hartung et al, "Cre mutants with altered DNA binding properties." J Biol Chem 1998; 273:22884-22891; Shaikh et al, "Chimeras of the Flp and Cre recombinases: Tests of the mode of cleavage by Flp and Cre. J Mol Biol. 2000; 302:27-48; Rongrong et al, "Effect of deletion mutation on the recombination activity of Cre recombinase." Acta Biochim Pol.

The term "recombine," or "recombination," in the context of a nucleic acid modification (e.g., a genomic modification), is used to refer to the process by which two or more nucleic acid molecules, or two or more regions of a single nucleic acid molecule, are modified by the action of a recombinase protein (e.g., an inventive recombinase fusion protein provided herein). Recombination can result in, inter alia, the insertion, inversion, excision, or translocation of nucleic acids, e.g., in or between one or more nucleic acid molecules.

The term "site-specific enzyme," as used herein, refers to any enzyme capable of binding a nucleic acid at a target site to mediate a modification of the nucleic acid. Typically, the site-specific enzymes provided herein comprise an intein (e.g., a ligand-dependent intein). In some embodiments, the site-specific enzyme is unable to bind a target site prior to excision of the intein. In some embodiments, the site-specific enzyme is able to bind a target site prior to excision of the intein but remains enzymatically inactive (e.g., cannot cleave, recombine, edit, or otherwise modify a nucleic acid) until excision of the intein.

The term "small molecule," as used herein, refers to a non-peptidic, non-oligomeric organic compound either prepared in the laboratory or found in nature. Small molecules, as used herein, can refer to compounds that are "natural product-like", however, the term "small molecule" is not limited to "natural product-like" compounds. Rather, a small molecule is typically a non-polymeric, non-oligomeric molecule that is characterized in
that it contains several carbon-carbon bonds, and has a molecular weight of less than 2000g/mol, preferably less than 1500g/mol, although this characterization is not intended to be limiting for the purposes of the present invention. In certain embodiments, the ligand of a ligand-dependent inteins used in the present invention is a small molecule.

The term "subject," as used herein, refers to an individual organism. In some embodiments, the subject is a human. In some embodiments, the subject is a non-human mammal. In some embodiments, the subject is a non-human primate. In some embodiments, the subject is a rodent. In some embodiments, the subject is a sheep, a goat, a cattle, a cat, or a dog. In some embodiments, the subject is a vertebrate, an amphibian, a reptile, a fish, an insect, a fly, or a nematode. In some embodiments, the subject is a research animal. In some embodiments, the subject is genetically engineered, e.g., a genetically engineered non-human subject. The subject may be of either sex and at any stage of development.

The terms "target nucleic acid," and "target genome," as used herein in the context of nucleases, refer to a nucleic acid molecule or a genome, respectively, that comprises at least one target site of a given nuclease.

The term "target site," refers to a sequence within a nucleic acid molecule that is bound and (1) cleaved; (2) recombined; (3) edited; or (4) otherwise modified by a site-specific enzyme. In some embodiments, a target site refers to a "nuclease target site," which is a sequence within a nucleic acid molecule that is bound and cleaved by a nuclease. A target site may be single-stranded or double-stranded. In the context of RNA-guided (i.e., RNA-programmable) nucleases (e.g., a Cas9 protein, a Cas9 variant, fragments of Cas9 or fragments of Cas9 variants, etc.), a target site typically comprises a nucleotide sequence that is complementary to a gRNA of the RNA-guided nuclease, and a protospacer adjacent motif (PAM) at the 3’ end adjacent to the gRNA-complementary sequence. For the RNA-guided nuclease Cas9, the target site may be, in some embodiments, 20 base pairs plus a 3 base pair PAM (e.g., NNN, wherein N represents any nucleotide). Typically, the first nucleotide of a PAM can be any nucleotide, while the two downstream nucleotides are specified depending on the specific RNA-guided nuclease. Exemplary target sites for RNA-guided nucleases, such as Cas9, are known to those of skill in the art and include, without limitation, NNG, NGN, NAG, and NGG, wherein N represents any nucleotide. In addition, Cas9 nucleases from different species (e.g., S. thermophilus instead of S. pyogenes) recognize a PAM that comprises the sequence: NGGNG. Additional PAM sequences are known, including, but not limited to, NNAGAAW and NAAR (see, e.g., Esvelt and Wang, Molecular Systems Biology, 9:641 (2013), the entire contents of which are incorporated herein by reference). For
example, the target site of an RNA-guided nuclease, such as, e.g., Cas9, may comprise the structure \([Nz]-[PAM]\), where each \(N\) is, independently, any nucleotide, and \(z\) is an integer between 1 and 50. In some embodiments, \(z\) is at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, or at least 50. In some embodiments, \(z\) is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50. In some embodiments, \(Z\) is 20. In some embodiments, "target site" may also refer to a sequence within a nucleic acid molecule that is bound but not cleaved by a nuclease.

[0066] The terms "transcriptional activator" and "transcriptional repressor," refer to agents which activate and repress the transcription of a gene, respectively. Typically, such activators and repressors are proteins, e.g., as provided herein.

[0067] The terms "treatment," "treat," and "treating," refer to a clinical intervention aimed to reverse, alleviate, delay the onset of, or inhibit the progress of a disease or disorder, or one or more symptoms thereof, as described herein. As used herein, the terms "treatment," "treat," and "treating" refer to a clinical intervention aimed to reverse, alleviate, delay the onset of, or inhibit the progress of a disease or disorder, or one or more symptoms thereof, as described herein. In some embodiments, treatment may be administered after one or more symptoms have developed and/or after a disease has been diagnosed. In some embodiments, treatment may be administered in the absence of symptoms, e.g., to prevent or delay onset of a symptom or inhibit onset or progression of a disease. For example, treatment may be administered to a susceptible individual prior to the onset of symptoms (e.g., in light of a history of symptoms and/or in light of genetic or other susceptibility factors). Treatment may also be continued after symptoms have resolved, for example, to prevent or delay their recurrence.

[0068] The term "vector" refers to a polynucleotide comprising one or more recombinant polynucleotides of the present invention, e.g., those encoding or a Cas9 protein (e.g., a Cas9 protein comprising an intein) and/or a gRNA provided herein. Vectors include, but are not limited to, plasmids, viral vectors, cosmids, artificial chromosomes, and phagemids. The vector is one which is able to replicate in a host cell, and which is further characterized by one or more endonuclease restriction sites at which the vector may be cut and into which a desired nucleic acid sequence may be inserted. Vectors may contain one or more marker sequences suitable for use in the identification and/or selection of cells which
have or have not been transformed or genomically modified with the vector. Markers include, for example, genes encoding proteins which increase or decrease either resistance or sensitivity to antibiotics (e.g., kanamycin, ampicillin) or other compounds, genes which encode enzymes whose activities are detectable by standard assays known in the art (e.g., β-galactosidase, alkaline phosphatase or luciferase), and genes which visibly affect the phenotype of transformed or transfected cells, hosts, colonies, or plaques. Any vector suitable for the transformation of a host cell, (e.g., E. coli, mammalian cells such as CHO cell, insect cells, etc.) as embraced by the present invention, for example vectors belonging to the pUC series, pGEM series, pET series, pBAD series, pTET series, or pGEX series. In some embodiments, the vector is suitable for transforming a host cell for recombinant protein production. Methods for selecting and engineering vectors and host cells for expressing gRNAs and/or proteins (e.g., those provided herein), transforming cells, and expressing/purifying recombinant proteins are well known in the art, and are provided by, for example, Green and Sambrook, Molecular Cloning: A Laboratory Manual (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)).

DETAILED DESCRIPTION OF CERTAIN EMBODIMENTS OF THE INVENTION

Site-specific enzymes which catalyze nucleic acid modifications are powerful tools for targeted genome modification in vitro and in vivo. Some site-specific enzymes can theoretically achieve a level of specificity for a target site that would allow one to target a single unique site in a genome for modification without affecting any other genomic site. In the case of site-specific nuclease, it has been reported that nuclease cleavage in living cells triggers a DNA repair mechanism that frequently results in a modification of the cleaved and repaired genomic sequence, for example, via homologous recombination or non-homologous end-joining. Accordingly, the targeted cleavage of a specific unique sequence within a genome opens up new avenues for gene targeting and gene modification in living cells, including cells that are hard to manipulate with conventional gene targeting methods, such as many human somatic cells or embryonic stem cells. Nuclease-mediated modification of disease-related sequences, e.g., the CCR-5 allele in HIV/AIDS patients, or of genes necessary for tumor neovascularization, can be used in the clinical context, and two site-specific nuclease are currently in clinical trials (Perez, E.E. et al., "Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nuclease." Nature Biotechnology. 26, 808-816 (2008); ClinicalTrials.gov identifiers: NCT00842634, NCT01044654, NCT01252641, NCT01082926). Other diseases that can be treated using
site-specific nucleases or other site-specific DNA modifying enzymes include, for example, diseases associated with triplet expansion (e.g., Huntington's disease, myotonic dystrophy, spinal-cerebellar ataxias, etc.), cystic fibrosis (by targeting the CFTR gene), cancer, autoimmune diseases, and viral infections.

[0070] One important problem with site-specific modification is off-target effects, e.g., the modification of genomic sequences that differ from the intended target sequence by one or more nucleotides. Undesired side effects of off-target modification range from insertion into unwanted loci during a gene targeting event to severe complications in a clinical scenario. For example, off-target modification of sequences encoding essential gene functions or tumor suppressor genes may result in disease or even the death of a subject. Accordingly, it is desirable to employ new strategies in designing site-specific enzymes having the greatest chance of minimizing off-target effects.

[0071] The systems, methods, and compositions of the present disclosure represent, in some aspects, an improvement over previous methods and compositions by providing means to control the spatiotemporal activity of site-specific enzymes, for example, RNA-guided nucleases and engineered RNA-guided nucleic acid modifying enzymes. For example, RNA-guided nucleases known in the art, both naturally occurring and those engineered, typically bind to and cleave DNA upon forming a complex with an RNA (e.g., a gRNA) that complements the target. Aspects of the present invention relate to the recognition that having spatiotemporal control of the enzymatic or nucleic acid binding properties of an RNA-guided nuclease and RNA-guided nucleic acid modifying enzymes by engineering variants to include an intein will decrease the likelihood of off-target effects by minimizing or controlling the time a RNA-guided nuclease or engineered RNA-guided nucleic acid modifying enzymes is active. Accordingly, the strategies, methods, compositions, kits, and systems provided herein can be used to control the activity of any site-specific enzyme (both naturally occurring and those engineered) such as RNA-guided nucleases (e.g., Cas9, Cas9 variants, fragments of Cas9 or Cas9 variants, etc.) or engineered nucleic acid modifying enzymes comprising a variant of an RNA-guided nuclease (e.g., dCas9).

[0072] Inteins are protein splicing elements that are able to catalyze their excision out of a single polypeptide and leave behind the flanking sequences, or exteins, precisely ligated together through a native peptide bond. Inteins are attractive tools for modulating protein structure and function because they do not require any other cellular components, are able to splice out of a wide variety of extein contexts, and can undergo splicing in minutes. Although natural inteins splice spontaneously, inteins that undergo splicing in a small
molecule-dependent or ligand-dependent manner have been developed by fusing intein halves with proteins that dimerize in the presence of a small molecule, or by directed evolution in which a library of intact inteins fused to a ligand-binding domain was screened to splice in the presence, but not the absence, of a small molecule or ligand. These ligand-dependent inteins have enabled protein function in cells to be controlled post-translationally by the addition of an exogenous, cell-permeable molecule (See e.g., published U.S. Patent Application US 2014/0065711 Al, the entire contents of which are hereby incorporated by reference). The inventors have found that the targeted insertion of ligand-dependent inteins into site-specific enzymes renders the enzymes, in some instances, inactive prior to the controlled excision of the intein through binding of a ligand specific for the intein. For example, the targeted insertion of a ligand-dependent intein into Cas9 at fifteen different positions resulted in a subset of Cas9 variants that were inactive in the absence of ligand, but upon addition of the ligand the intein self-excised resulting in an active Cas9 protein capable of site-specific cleavage of a target gene.

Some aspects of this disclosure are based on the surprising discovery that Cas9 proteins comprising an intein, for example, a ligand-dependent intein as described herein, exhibit an increased specificity as compared to constitutively active Cas9 proteins. For example, it was found that the conditionally active Cas9 proteins comprising an intein exhibit an activity in the "on" state that is comparable to wild-type Cas9 activity or only slightly decreased as compared to wild-type Cas9 activity, while exhibiting decreased off-target activity.

In addition, some aspects of this disclosure relate to the recognition that Cas9 off-target activity is at least in part related to the concentration of active Cas9 proteins, and that the off-target activity of the provided conditionally active Cas9 proteins, e.g., the provided ligand-dependent Cas9 proteins, can be modulated, e.g., further decreased, by contacting the Cas9 proteins with a minimal amount of ligand effecting the desired result, e.g., the minimal amount effecting intein excision from a Cas9 protein, or the minimal amount resulting in a desired level of Cas9 protein activity.

While of particular relevance to DNA and DNA-cleaving nucleases such as Cas9 and variants thereof, the inventive concepts, methods, compositions, strategies, kits, and systems provided herein are not limited in this respect, but can be applied to any nuclease or nucleic acid:enzyme system utilizing nucleic acid templates such as RNA to direct binding to a target nucleic acid. For example, the inventive concepts provided herein can be applied to RNA-guided nucleic acid-targeting protein, e.g., to RNA-guided nucleases, and to fusion
proteins comprising nucleic acid-targeting domains of such nucleases, e.g., to fusion proteins comprising a Cas9 targeting domain (e.g., dCas9 domain), and a functional (effector) domain, such as, for example, a heterologous nuclease domain, recombinase domain, or other nucleic acid-editing domain.

Small molecule controlled site-specific enzymes

Some aspects of this disclosure provide site-specific enzymes engineered to have both an "on" and "off" state which depends on the presence of a ligand such as a small molecule. The ligand binds and activates the enzyme through binding a ligand-dependent intein in the enzyme, whereby ligand binding induces self-excision of the intein thereby activating the enzyme (e.g., the presence of the intein in the enzyme disrupted one or more activities of the enzyme). In some aspects then, the enzymes may collectively be referred to as "small molecule controlled" or "ligand-dependent" site-specific enzymes. In some embodiments, the site-specific enzyme that has been modified to include a ligand-dependent intein comprises Cas9, or a variant of Cas9.

Accordingly, in the absence of a ligand that binds the intein, the intein is not excised, and the protein comprising Cas9 or variant of Cas9 remains inactive. By "inactive" it is meant that the protein has no or minimal activity with respect to one or more activities described herein. In some embodiments, prior to intein excision, the protein has (i) no or minimal enzymatic activity; (ii) no or minimal gRNA binding activity; (iii) no or minimal target nucleic acid binding activity; or any combination of (i)-(iii), e.g., the protein has (i) and (ii); (i) and (iii); (ii) and (iii); or (i), (ii) and (iii). Enzymatic activities for (i), include, for example, nuclease activity, nickase activity, recombinase activity, nucleic acid editing (e.g., deaminase) activity, transcriptional activation, transcriptional repression, and epigenetic modification activity.

In some embodiments, by "minimal" activity, it is meant that the protein, prior to excision of the intein, exhibits less than 50%, less than 45%, less than 40%, less than 35%, less than 30%, less than 25%, less than 24%, less than 23%, less than 22%, less than 21%, less than 20, less than 19%, less than 18%, less than 17%, less than 16%, less than 15%, less than 14%, less than 13%, less than 12%, less than 11%, less than 10%, less than 9%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, or less than 1% of a particular activity (e.g., nuclease activity, nickase activity, recombinase activity, deaminase activity, transcriptional activation, transcriptional repression, epigenetic modification activity, gRNA binding activity, and/or target nucleic acid binding activity) as
compared to either the wild type counterpart of the protein or the intein-excised form of the protein. In some embodiments, following excision of the intein, the protein exhibits at least a 1.25-fold increase, at least a 1.5-fold increase, at least a 1.75-fold increase, at least a 2.0-fold increase, at least a 2.25-fold increase, at least a 2.5-fold increase, at least a 2.75-fold increase, at least a 3.0-fold increase, at least a 3.25-fold increase, at least a 3.5-fold increase, at least a 3.75-fold increase, at least a 4.0-fold increase, at least a 4.5-fold increase, at least a 5.0-fold increase, at least a 5.5-fold increase, at least a 6.0-fold increase, at least a 6.5-fold increase, at least a 7.0-fold increase, at least a 7.5-fold increase, at least a 8.0-fold increase, at least a 8.5-fold increase, at least a 9.0-fold increase, at least a 9.5-fold increase, or at least a 10.0-fold or more increase in activity (e.g., nuclease activity, nickase activity, recombinase activity, or deaminase activity) as compared to the intein-intact form of the protein. Methods for assessing the activity of any ligand-dependent site-specific Cas9-containing enzyme provided herein are well known to those of ordinary skill in the art, and in the context of nuclease activity include those described in the Examples.

In some embodiments, upon excision, the intein leaves a cysteine residue. Thus, if the intein is inserted such that it replaces a cysteine, the Cas9 protein, upon intein excision, will be unmodified as compared to the original protein. If the intein replaces any other amino acid, the Cas9 protein, upon intein excision, will contain a cysteine in place of the amino acid that was replaced. In some embodiments, the intein does not replace an amino acid residue in a Cas9 protein, but is inserted into the Cas9 protein (e.g., in addition to the amino acid residues of the Cas9 protein). In this aspect, upon excision, the protein will comprise an additional cysteine residue. While the presence of an additional cysteine residue (or the substitution of a residue for a cysteine upon excision) is unlikely to affect the function of the Cas9 protein, in some embodiments where the intein does not replace a cysteine, the intein replaces an alanine, serine, or threonine amino acid, as these residues are similar in size and/or polarity to cysteine.

Accordingly, in some embodiments, the intein is inserted into one or both of the nuclease domains of Cas9 or a Cas9 variant (e.g., dCas9, Cas9 nickase), such as the HNH domain and/or the RuvC domain. In some embodiments, the intein is inserted into one or more other domains of Cas9 or a Cas9 variant (e.g., dCas9, Cas9 nickase), such as, REC1, REC2, PAM-interacting (PI), and/or bridge helix (BH) domain. The sequences and structure corresponding to these domains are known, and in some aspects are represented by the underlined segments of SEQ ID NO:2 (Cas9) and SEQ ID NO:5 (dCas9) above (See also, Nishimasu et al., "Crystal structure of Cas9 in complex with guide RNA and target DNA."
In some embodiments, the intein is inserted into any location of Cas9, e.g., any location that disrupts one or more activities of Cas9 (e.g., enzymatic activity, gRNA binding activity, and/or DNA binding activity). In some embodiments, the intein is inserted into a sequence of Cas9 or a Cas9 variant such that the intein sequence replaces one or more amino acids in the protein. In some embodiments, the intein replaces any cysteine, any alanine, any threonine, or any serine residue in Cas9 or a Cas9 variant including Cas9 nickase and dCas9 (and fusions thereof). In some embodiments the inserted intein sequence replaces Cys80, Ala127, Thr46, Ser219, Thr333, Thr519, Cys574, Thr622, Ser701, Ala728, Thr995, Ser1006, Ser1154, Ser1159, or Ser1274 in Cas9 (SEQ ID NO:2), dCas9 (SEQ ID NO:5), or Cas9 nickase (SEQ ID NO:4). In some embodiments, the intein is inserted within 5, within 10, within 15, or within 20 amino acids of Cys80, Ala127, Thr46, Ser219, Thr333, Thr519, Cys574, Thr622, Ser701, Ala728, Thr995, Ser1006, Ser1154, Ser1159, or Ser1274 in Cas9 (SEQ ID NO:2), dCas9 (SEQ ID NO:5), or Cas9 nickase (SEQ ID NO:4). In some embodiments, the inserted intein sequence replaces Ala27, Thr46, Ser219, Thr519, or Cys574 in Cas9 (SEQ ID NO:2), dCas9 (SEQ ID NO:5), or Cas9 nickase (SEQ ID NO:4). In some embodiments, a Cas9 protein comprising an intein comprises an amino acid sequence selected from the group consisting of SEQ ID NOs:27-41, or comprises an amino acid sequence that has at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% sequence identity to any one of SEQ ID NOs:27-41. In some embodiments, the intein is inserted into the protein such that it does not replace any amino acid, but is added in addition to the amino acid sequence of the protein. The intein that is inserted into the protein can be any ligand-dependent intein, e.g., those described herein. For example, in some embodiments, the intein that is inserted into the protein comprises, in part or in whole, a sequence that is at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% identical to any one of SEQ ID NO:27-41.

Cas9:Intein (37R3-2; in double underline) replacing Cys80

Cas9: Intein (37R3-2; in double underline) replacing AlaL27

MDKKS IGDLGTVSNGWAVIDEYKPSKFKVLGNTDRHS IKNLIGALLFDGETEAELRTKRTA
RRYTRRNKRICYLCQIEISFENAKVDSFHPKLRLEESLVDEKKHREIFPGNI VDEVCIAEGTRIFD
PTVPGTHIRDVSDKPIHVYAMAKDGLTALLERVWSFDQGTRDVIGLRIAGGAI WATPDYKILTE
YGWRASAEKGLKRDVGAPGSNSGLSLTAQDMVSALLAAPLYEEDTSPFSEASMMGLTLNL
ADRELHVMINAWRKPGVFVLDLHQLQHQAHLERRAEWILMGLVWRSMEHPGFKLFLAPFNLLDLRNNQCGK
VEGMEVEIFDMLLSSATNRMNQLLQEEFVCLKIS LLLNSQVTFLSSTLSEKHDHIALDKITDL
LHLMKAKGLTLIQOQPQLRQALLLSIRHLSNKGMHKSMYSHKVNVPVLYDLLEMDALHRLLGGG
SGASRQVAFADALDDKLHMDAELGRYSLVIREVLPRTRARTDFDLEVEELHTLVAE GVVVHNCYHEKY
PTHYLHRKLVSTDDLRLYLALYHAKIFGRHLFGLDGNDSVDKLFQIVTNQLFEENP
INASGEKJASLRLSKRSRELNI AQLPGKKNFLGFGNI ALSLGLTPFNKFSDLADEALKLQSK
DTYDDDLNINOLAQIDQYADLFAAKNSLADILSDLIRVNTETIKAPLSAMSIKRYDEHDQOLTTK
ALVRQQLPKEKYKEIFFDQSFKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT
FDNGSIPHQIHLGELHAILRRQIECDFSVEI SGEDRFNASLGTYHDLKI IKKDFLFD
NEENILEDIVDITLTLFLEDREMIERKTAYAHLDDKVQKLRRTYWGRSLKLIRGDKQSG
KTILDWLDGSFHKPENIEMARENQTTQGKQNSRMRKREIEQIKELGSQIKHEFVENTLQNEKY
LLYLQGMRDQMDVQEDINLSDYDVFQHPSFLKDSIDNKTVLSRDKNKQSDNVPSEEVVKKMK
NYWQLNNLAKLITQRDFDNLRRGKSGHEJAFKIRQLVETRQITKHVALQLDAMDNTKYENDK
LIREVKTILSKKLVSDFRDKFQYKVRLEINNVHHHAYDAMLNAVGTALKYKPLESEFVYGDKYV
DVRKMI AKSEQEIQKATAYFFYSNIMNFKTEITLANGEIRKPIETNGEII WVDGRDFATVR
KVLSPQVNI VKKTEVQTFGQESLKPANSKDLI ARKKDWFPKGGFDSTVAYSLVVLAVKVEKG
KSKKLKSVKELLGASRQALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSRKVLADANLD
KVLASYNKHRDKFIREQAENI IHLFTLNLGAPAAFPFYDFTIDDRKYSTKEVLATLHQSITGLY
ETRIDLSQGDD (SEQ ID NO: 27)

Cas9: Intein (37R3-2; in double underline) replacing ThrL46

MDKKS IGDLGTVSNGWAVIDEYKPSKFKVLGNTDRHS IKNLIGALLFDGETEAELRTKRTA
RRYTRRNKRICYLCQIEISFENAKVDSFHPKLRLEESLVDEKKHREIFPGNI VDEVAHYEKYPTI
HLRKLVDSCIALEGTRIFDPVTGTHIRDVSDKPIHVYAMAKDGLTALLERVWSFDQGTRDVIGL
RIAGGAIWATPDHKLTEYGWRAAELRIKGDVFAGPSNSGLSLTAQDMVSALLADAPLPYEEDTSPFSEASMMGLTLNL
ADRELHVMINAWRKPGVFVLDLHQLQHQAHLERRAEWILMGLVWRSMEHPGFKLFLAPFNLLDLRNNQCGK
VEGMEVEIFDMLLSSATNRMNQLLQEEFVCLKIS LLLNSQVTFLSSTLSEKHDHIALDKITDL
LHLMKAKGLTLIQOQPQLRQALLLSIRHLSNKGMHKSMYSHKVNVPVLYDLLEMDALHRLLGGG
SGASRQVAFADALDDKLHMDAELGRYSLVIREVLPRTRARTDFDLEVEELHTLVAE GVVVHNCYHEKY
PTHYLHRKLVSTDDLRLYLALYHAKIFGRHLFGLDGNDSVDKLFQIVTNQLFEENP
INASGEKJASLRLSKRSRELNI AQLPGKKNFLGFGNI ALSLGLTPFNKFSDLADEALKLQSK
DTYDDDLNINOLAQIDQYADLFAAKNSLADILSDLIRVNTETIKAPLSAMSIKRYDEHDQOLTTK
ALVRQQLPKEKYKEIFFDQSFKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT
FDNGSIPHQIHLGELHAILRRQIECDFSVEI SGEDRFNASLGTYHDLKI IKKDFLFD
NEENILEDIVDITLTLFLEDREMIERKTAYAHLDDKVQKLRRTYWGRSLKLIRGDKQSG
KTILDWLDGSFHKPENIEMARENQTTQGKQNSRMRKREIEQIKELGSQIKHEFVENTLQNEKY
LLYLQGMRDQMDVQEDINLSDYDVFQHPSFLKDSIDNKTVLSRDKNKQSDNVPSEEVVKKMK
NYWQLNNLAKLITQRDFDNLRRGKSGHEJAFKIRQLVETRQITKHVALQLDAMDNTKYENDK
LIREVKTILSKKLVSDFRDKFQYKVRLEINNVHHHAYDAMLNAVGTALKYKPLESEFVYGDKYV
DVRKMI AKSEQEIQKATAYFFYSNIMNFKTEITLANGEIRKPIETNGEII WVDGRDFATVR
KVLSPQVNI VKKTEVQTFGQESLKPANSKDLI ARKKDWFPKGGFDSTVAYSLVVLAVKVEKG
KSKKLKSVKELLGASRQALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSRKVLADANLD
KVLASYNKHRDKFIREQAENI IHLFTLNLGAPAAFPFYDFTIDDRKYSTKEVLATLHQSITGLY
ETRIDLSQGDD (SEQ ID NO: 28)
PGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLATSSRFRMMNLQGEEFVCLKS I I L N S G V Y T F L S T T L
KSLEEKDHIIARLKDITDLHMAKAGLTLOQQHQHRLAQLLLILSIRHMSNKGMHLYSMKYKNVV
FLYDILLEMLDAHRHLHAGGSGASRVQAFADALDDKFLHDLMAEGLRYSVIREVLTPRTRAPDLEVEE
LHTLVAEVGVVHNC DKADLRLI YLalahMkIFRQHFLIEGDLNPDNSDVKLFLQVTONQLFEEENP
INASGVDAKAILARLSKSRRELNI AQLPGPKGNaGLFGNLINALSGLTPFNKSFLAEADKQLKS
DITYDDDLNLQAIDQYATIAAKNLSDAIISSDLDVNEITIKAPLASIKMRHDEHQDILTTLK
ALVQRQLPEKYKEIFFQSKNGYAGYIDGGASQEEFYKFIKPLEMDGTEELLVKnREDLRRQRT
FDNGSIPHQIALGELHAIIRQEDFYPFLKDNREIKIILTFRIPYYVQPLGARNSRFAMTRKSEET
ITPWNFEVEVDKGASQFIERMTNDFKLNPEKVLPSHLYEFTYVENLTQKYVTEGMRPFAFL
SGEQKKAI DVLFLFKTNKSVTVKQLKEDFYKIECFDSVEI SGVEDRFNASLGTYHDLKLI IKDKDI
NEENEDILEDIVTLTLEDFDYLKQYHLDFFKPRGWTGGLRLINWRLKRDQSG
KTLDFKLSDGFANRNFMQLIHDDSLTALKEDIQKAQVSGQGDSLHEHI ANLAPSaIAKKGILQTQVV
DELVKGVRHKPENEIEMARENQTQTKDQKNSSERNKRIEGEKIELGSQILKEHPVENTIQNLEKYL
LILQNGRDMPVYDQSGAIINLTLNADRELVPQMSKLDSDINVIKLQRSNQGKSDNMSEEVVKMK
NYWRQLNAKVTQFDNLTKERGGSLSEDKAGFJKRQLSVERTITQKHVAIILDSRMNTKENDK
LISREVITLKLVSILVKEIQLDKFQFYQKVERINNYHQAAYAQAADLVATLXILYKPLESFVQDYKKV
DVRKMI AASEQEGIEQKATAYFFYNIMNFKTEITLANGEIRKPLIETNGETGIE VDWKGRDFATVRR
KVLSPMQVI VVKTETEVQNLKEPSKILPSRNVEKLI AKRPWDPPKYYGFDSPFVASYLVVAKEK
SKKLLKSVKELLI TIMERS SPEKNPFDLEAKGYVEKVKDI IILKPYSLEFENGKRLMASAGEL
QKGNELALPSKLYASHKLGSFEDNEQKLFVQHEQKHYLDEI IEQQIEFSKVRILANADL
KVLASYKHKRDKPIREQAENI IHLFTLTNLGAPAAFKYFDDTTIDRKYSTKELVATLHIQSTGLY
ETRIDLSQLGGD (SEQ ID NO:29)

Cas9:Intein (37R3-2; in double underline) replacing Ser219
MDKYS IGLDGTNGWVAPIDTEYKVPSSKFKVLQNGDHS I KKNLIALLFDGETEAEATRLKRTA
RRRTRYRKNCRLCYQIEFSNEMAKVVDFFSRHELLEESVVEEEDKHERHIFPGNI VDEVAHYEKPTIY
HLRRKLVSTDIAKDLRLI YLalahMkIFRQHFLIEGDLNPDNSDVKLFLQVTONQLFEEENP INAS
GVDAKAILARLSKCLAEGRTRIPFDVPVTGTHRILEDVQGDPKIHPHVAAAKDGLLRAPVWSFDQTR
DVGILRIAGAI VWATPDHKVLTEYWGRAWAGELRKGDRVAGPSGGLNSLAALSITADQMSALLDAEPP
ILYSEYDTPSFSEASGMINGLTLNADRELVPQMSKLDSDINVIKLQRSNQGKSDNMSEEVVKMK
NYWRQLNAKVTQFDNLTKERGGSLSEDKAGFJKRQLSVERTITQKHVAIILDSRMNTKENDK
LISREVITLKLVSILVKEIQLDKFQFYQKVERINNYHQAAYAQAADLVATLXILYKPLESFVQDYKKV
DVRKMI AASEQEGIEQKATAYFFYNIMNFKTEITLANGEIRKPLIETNGETGIE VDWKGRDFATVRR
KVLSPMQVI VVKTETEVQNLKEPSKILPSRNVEKLI AKRPWDPPKYYGFDSPFVASYLVVAKEK
SKKLLKSVKELLI TIMERS SPEKNPFDLEAKGYVEKVKDI IILKPYSLEFENGKRLMASAGEL
QKGNELALPSKLYASHKLGSFEDNEQKLFVQHEQKHYLDEI IEQQIEFSKVRILANADL
KVLASYKHKRDKPIREQAENI IHLFTLTNLGAPAAFKYFDDTTIDRKYSTKELVATLHIQSTGLY
ETRIDLSQLGGD (SEQ ID NO:30)
Cas9:Intein (37R3-2; in double underline) replacing Thr333

MDKKYS IGLDIGTQISWVFEKPVKFKVLGNTDIVDGLGTYSTLQHILNLKQSDVQGSLQSDSDDHQLDGGKMDK
MDKKYS IGLDIGTQISWVFEKPVKFKVLGNTDIVDGLGTYSTLQHILNLKQSDVQGSLQSDSDDHQLDGGKMDK
QKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI IEQISEFSKRVILADANL D KVLASYNHKRDKRQFIRQAQE HILFELTNAGAPAFAFKFYDIDDRKRYTSTKEVLDATLIHQ SITGLY ETRIDLSQQLGDD (SEQ ID NO:32)

Cas9: Intein (37R-2; in double underline) replacing Cys574

MDKKYS IGLDIGTNSGWAVITDEYKVPSSKKFKVLGNTDRHS IKKNLIGALLFDSGETAEATRLKRTA RRYTTRKNRICYLQIEFSNEMAKVDSDSFHRLEFLVEEDEKKHERPHFGNI VDEVAYHEKPYTI HLRRKLVSSTDKDALLLI YLLAHMKIFRQHFLIEGDLNPDNSVDKFIQLVQTYNLFEEPINAS GVDAAKIALSRLSKSRLENLI AQLPGKNGKLFGLNLI AGLSLGTPFNSFNLAEADAKIQLSKDTY DDDNLQAIGQDYADLPLAAKNSDAILSDILDLVNEITEKAPLSAMIKRYDEHHQDLTLKLALVR QOLPEKYKIEFDQQNGYQHYTIDGQSEEFYKFLKPILEKMDGTELVKDLNREDLRKRQFTDNG SIPHQILGHGELLIRQDEPFYPLKDNRKIEKILIFHRIPYVYPGLRAGNSRFAWMTKSEETITPW NFEENVKGAQSFQFERTMNFKNLEKPVHSILYYFETYVNETLTKYKTEGMRKPAFLSSEQ KKAIVDILLFKTRNKTVVQKVLKDEEPIFRDPTVTTHIREDVHDRGKPIHVVAAADGTL LLLARFVSFWDOGQTDVGILRI AGGAIWATPDHKVLTEYGWRAAGELKGRDVAQPGGNSINASELSL TADQMSVALLDAAEPPILYSEYDPTSPFSEAMSGMLLTNLADRLHIWUMNRAVPGFVDLTLHDAQHL \nLERAWLEILMIGLWRSMEHPPKLLFAAPNPLLDRNOQKCVEGMVEIFDMILLASSFRMNNQEEFV CLKSIIILLNSGVYFTLSSTLKSLEEDHIDHRAKDIITDLHLMAKAGLTLQQHQHRLAQLLLHLS \ RHMNSNKMEHLSYVLDDMLDDALHRHASSQVRAGFAVADLKLFDHDEGLREYRS IEMARENQTTQKQNSRCRMIIEGIELSqliKEHPVENTQLNEKLY LYYLYQGRMDVQGQEDNLYYIDGQHSEEFYQFSLKDIDSNKTVLSDNRKGSNDFSEVVKKMK NYRQWMLNACLITQRFDNLTAKKGELSDKAGFKIQRLQETIQKVAIQLDSRTNKYDENK LIREKVVTILKSKDFRSDKFRKVREINYYHAHADLYNAVTGILALKKPLESEFVYDGKYV DVRKMI AKESEQIEKATAKYFVSNIMNFMKTETITLANEIKRRLPIEITNETGEI WDVGKGRDFATVR KVLSPMQVNI VKKTEVQTGQGSKESILPKRNSDKL IAKRDWDPKYYYGGDFSDPTVAYSVLVVAVERG KSKKLKSVKELLGI TIMERS SFEKF FDEAKYKEVKDL IIEKFKYSLELENGRKMLASAGEL QKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI IEQISEFSKRVILADANL D KVLASYNHKRDKRQFIRQAQE HILFELTNAGAPAFAFKFYDIDDRKRYTSTKEVLDATLIHQ SITGLY ETRIDLSQQLGDD (SEQ ID NO:33)

Cas9: Intein (37R-2; in double underline) replacing Thr622

MDKKYS IGLDIGTNSGWAVITDEYKVPSSKKFKVLGNTDRHS IKKNLIGALLFDSGETAEATRLKRTA RRYTTRKNRICYLQIEFSNEMAKVDSDSFHRLEFLVEEDEKKHERPHFGNI VDEVAYHEKPYTI HLRRKLVSSTDKDALLLI YLLAHMKIFRQHFLIEGDLNPDNSVDKFIQLVQTYNLFEEPINAS GVDAAKIALSRLSKSRLENLI AQLPGKNGKLFGLNLI AGLSLGTPFNSFNLAEADAKIQLSKDTY DDDNLQAIGQDYADLPLAAKNSDAILSDILDLVNEITEKAPLSAMIKRYDEHHQDLTLKLALVR QOLPEKYKIEFDQQNGYQHYTIDGQSEEFYKFLKPILEKMDGTELVKDLNREDLRKRQFTDNG SIPHQILGHGELLIRQDEPFYPLKDNRKIEKILIFHRIPYVYPGLRAGNSRFAWMTKSEETITPW NFEENVKGAQSFQFERTMNFKNLEKPVHSILYYFETYVNETLTKYKTEGMRKPAFLSSEQ KKAIVDILLFKTRNKTVVQKVLKDEEPIFRDPTVTTHIREDVHDRGKPIHVVAAADGTL LLLARFVSFWDOGQTDVGILRI AGGAIWATPDHKVLTEYGWRAAGELKGRDVAQPGGNSINASELSL TADQMSVALLDAAEPPILYSEYDPTSPFSEAMSGMLLTNLADRLHIWUMNRAVPGFVDLTLHDAQHL \nLERAWLEILMIGLWRSMEHPPKLLFAAPNPLLDRNOQKCVEGMVEIFDMILLASSFRMNNQEEFV CLKSIIILLNSGVYFTLSSTLKSLEEDHIDHRAKDIITDLHLMAKAGLTLQQHQHRLAQLLLHLS \ RHMNSNKMEHLSYVLDDMLDDALHRHASSQVRAGFAVADLKLFDHDEGLREYRS IEMARENQTTQKQNSRCRMIIEGIELSqliKEHPVENTQLNEKLY LYYLYQGRMDVQGQEDNLYYIDGQHSEEFYQFSLKDIDSNKTVLSDNRKGSNDFSEVVKKMK NYRQWMLNACLITQRFDNLTAKKGELSDKAGFKIQRLQETIQKVAIQLDSRTNKYDENK LIREKVVTILKSKDFRSDKFRKVREINYYHAHADLYNAVTGILALKKPLESEFVYDGKYV DVRKMI AKESEQIEKATAKYFVSNIMNFMKTETITLANEIKRRLPIEITNETGEI WDVGKGRDFATVR KVLSPMQVNI VKKTEVQTGQGSKESILPKRNSDKL IAKRDWDPKYYYGGDFSDPTVAYSVLVVAVERG KSKKLKSVKELLGI TIMERS SFEKF FDEAKYKEVKDL IIEKFKYSLELENGRKMLASAGEL QKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI IEQISEFSKRVILADANL D KVLASYNHKRDKRQFIRQAQE HILFELTNAGAPAFAFKFYDIDDRKRYTSTKEVLDATLIHQ SITGLY ETRIDLSQQLGDD (SEQ ID NO:33)
Cas9: Intein (37R3-2; in double underline) replacing Ser701

MDKKYS IGLDINTGNSVGWAVITDEYKPSKFKVLGVNGTDHR5 IKKNLIGALLFDGSETAEATRLKRTA
RRRTDITIONCYLQEIFSNEMAKVDSSFHRLESSFLVEDDKHERHPFNGI VDEAYHEKYPTIY
HLRRKVLSTDKADLLRL YLALHAMIKFRGKFLIEGDLDNLPSDFDKDFQFYKVREINNYHHAHDAYLNAVGTALIKKYKPLESEFVYGDVKYVY
DVRRMKI AKSEQEQIGATA KYYFNSMFFKTEITLANGEIRKRPLFIETNETGETEI VWDGRDFATVR
KVLSPQVNI VKKTEVQTTGSFLPFKFRNSDLDK ARKKWDPPKYGGFDSPTAVSLVVAKVKEG
KSSKLKSVKELLIG TIMERS SPEKNFIPFDLQYEKAVKDDL IILKFKYSLENFERGKRMASAGEL
QGKNELALPSKYNVFYLASHYKELGSTFEDNEQKQLFVQKHYLDE IEIQISEFSKVRILADANL
KVLASYNHKRDPIREQAENI IILFHTLNLGAPAAFKYFDTIDIRKYSTKEVDATLHIQSTIGLY
ETRIDLSQQLGDD (SEQ ID NO:34)

Cas9: Intein (37R3-2; in double underline) replacing Ser701

MDKKYS IGLDINTGNSVGWAVITDEYKPSKFKVLGVNGTDHR5 IKKNLIGALLFDGSETAEATRLKRTA
RRRTDITIONCYLQEIFSNEMAKVDSSFHRLESSFLVEDDKHERHPFNGI VDEAYHEKYPTIY
HLRRKVLSTDKADLLRL YLALHAMIKFRGKFLIEGDLDNLPSDFDKDFQFYKVREINNYHHAHDAYLNAVGTALIKKYKPLESEFVYGDVKYVY
DVRRMKI AKSEQEQIGATA KYYFNSMFFKTEITLANGEIRKRPLFIETNETGETEI VWDGRDFATVR
KVLSPQVNI VKKTEVQTTGSFLPFKFRNSDLDK ARKKWDPPKYGGFDSPTAVSLVVAKVKEG
KSSKLKSVKELLIG TIMERS SPEKNFIPFDLQYEKAVKDDL IILKFKYSLENFERGKRMASAGEL
QGKNELALPSKYNVFYLASHYKELGSTFEDNEQKQLFVQKHYLDE IEIQISEFSKVRILADANL
KVLASYNHKRDPIREQAENI IILFHTLNLGAPAAFKYFDTIDIRKYSTKEVDATLHIQSTIGLY
ETRIDLSQQLGDD (SEQ ID NO:35)
LATSSRFMMNLQGEFFVCLKS I LNILSSGVYTFLSSLKSLSEEEK HDHRLDKTDTL IHLMAKAGLT LQQQHQQLAQ LLLLSHLISHHMNSNKMEHLYSMKYKNVVPYLYDLLLLMADHRLHAGGSGASVRQAFAD ALDDKFHLDMALRSVYRISVIEVLPRRTDLEVEELHTLVAEVGVHVHNCGSAIKKGLQTVV VDELKVMGRHKPENIEMARENQTTQKQGKNSRERMKRIEEGIKELGSLQIKHPVENTLQNEKLY LLYQLNDRMYQDDELINRLSVDVFVQSPKFLSDIDSNKVRSDKNKGSDFESEVVKAKM NYWQQLNACLITQKAKYFFYSINMMFFKETITALGERKPLSETGETGEI VWDGKRDFATVR KVLSPMQVNI VKKTEVTQQGTFKESLPLKRNSDKL ARKKDPKKYGGFSDPVTAYSVLVVAKVEKG KSKKLKSVKELLGI TIMERS SFEKNPDLFLAEKYGKVEKVDL I I KLPKYSLEFENGRKLMASAGEL QKGNELAPLSYVLASYKELKSGPEDNEQKQLFVEQHKHYLDEI IEIQISEFSKRVILADANL KVLASAYNHKRDPLFIREQAEI IHFLTTLNLGAPAFAKYFDTTIDRKYSTKEVLATLHIQS ITGLYPETRTSLQGGLD (SEQ ID NO:36)

Cas9:Intein (37R3-2; in double underline) replacing Thr995

MDKKS IGLDIGNSTSGVAVITDEKYPSKKFVKLGLNDRHS IKKNLIGALFDGSGETEAATRLKRTA RRRYTRRRKNICLQLRPSNEMAKVDSFDHRLKLESLEEDKHVRHFIGNI VDEAVAYHEKPYTI HLRRKLVSTDKADLLRI YALAHMKRIFGHFLIEGDLNPDNSVDKFIQLVQTVNLQFEENINAS GVDKAIALSRALKSRRLENLI CLQPEKQFQGNFLGI LANSITLPGFNSDFLADKLSQKDSTY DDDLNAAIQGQYDAPLFAANLSDAILDLSSDLVRVTKEAIPLASA MRKDEHHQQDTLLKLAVQR QLPEKYEKIFFKDFQFYKVREINNYHADAYLNAV VCKVITKLKSVLDSFRKDFQFYKVREINIYYHADAYLNAV CGAEGRDFPVTGHTIREVDGRKP KHVVAAAKDGTLLARPVVSFDOQIDTRVGI LRRAGGAI VVATPDHKVLEYGWRAAGEKLRGDRVAGPGSNSLALSITADQMVSALLDAEFPILILYSEYDPTSFSFAEASMMLNLADRELVMINWAKRP GFVDLTLHDOAHLLERAWEILIMGLWRSMEHGPKLFLAPNLLDLRRNQGKCVE GMVEIFDMLATTS RFRMNMLQGEFFVCLKS I LNILSSGVYTFLSSLKSLSEEEK HDHRLDKTDTL IHLMAKAGLT LQQQHQQLAQ LLLLSHLISHHMNSNKMEHLYSMKYKNVVPYLYDLLLLMADHRLHAGGSGASVRQAFAD ALDDKFHLDMALRSVYRISVIEVLPRRTDLEVEELHTLVAEVGVHVHNCGSAIKKGLQTVV VDELKVMGRHKPENIEMARENQTTQKQGKNSRERMKRIEEGIKELGSLQIKHPVENTLQNEKLY LLYQLNDRMYQDDELINRLSVDVFVQSPKFLSDIDSNKVRSDKNKGSDFESEVVKAKM NYWQQLNACLITQKAKYFFYSINMMFFKETITALGERKPLSETGETGEI VWDGKRDFATVR KVLSPMQVNI VKKTEVTQQGTFKESLPLKRNSDKL ARKKDPKKYGGFSDPVTAYSVLVVAKVEKG KSKKLKSVKELLGI TIMERS SFEKNPDLFLAEKYGKVEKVDL I I KLPKYSLEFENGRKLMASAGEL QKGNELAPLSYVLASYKELKSGPEDNEQKQLFVEQHKHYLDEI IEIQISEFSKRVILADANL KVLASAYNHKRDPLFIREQAEI IHFLTTLNLGAPAFAKYFDTTIDRKYSTKEVLATLHIQS ITGLYPETRTSLQGGLD (SEQ ID NO:37)

Cas9:Intein (37R3-2; in double underline) replacing Serl006

MDKKS IGLDIGNSTSGVAVITDEKYPSKKFVKLGLNDRHS IKKNLIGALFDGSGETEAATRLKRTA RRRYTRRRKNICLQLRPSNEMAKVDSFDHRLKLESLEEDKHVRHFIGNI VDEAVAYHEKPYTI HLRRKLVSTDKADLLRI YALAHMKRIFGHFLIEGDLNPDNSVDKFIQLVQTVNLQFEENINAS GVDKAIALSRALKSRRLENLI CLQPEKQFQGNFLGI LANSITLPGFNSDFLADKLSQKDSTY DDDLNAAIQGQYDAPLFAANLSDAILDLSSDLVRVTKEAIPLASA MRKDEHHQQDTLLKLAVQR QLPEKYEKIFFKDFQFYKVREINNYHADAYLNAV VCKVITKLKSVLDSFRKDFQFYKVREINNYHADAYLNAV CGAEGRDFPVTGHTIREVDGR KP KPIHVVAAAKDGTLLARPVVSFDOQIDTRVGI LRRAGGAI VVATPDHKVLEYGWRAAGEKLRGDRVAGPGSNSLALSITADQMVSALLDAEFPILILYSEYDPTSFSFAEASMMLNLADRELVMINWAKRP GFVDLTLHDOAHLLERAWEILIMGLWRSMEHGPKLFLAPNLLDLRRNQGKCVE GMVEIFDMLATTS RFRMNMLQGEFFVCLKS I LNILSSGVYTFLSSLKSLSEEEK HDHRLDKTDTL IHLMAKAGLT LQQQHQQLAQ LLLLSHLISHHMNSNKMEHLYSMKYKNVVPYLYDLLLLMADHRLHAGGSGASVRQAFAD ALDDKFHLDMALRSVYRISVIEVLPRRTDLEVEELHTLVAE GVVCVNNCALIKYPKLSEFEVYDGYKVK DVERMKIIQSEKQFQKETITALGERKPLSETGETGEI VWDGKRDFATVR KVLSPMQVNI VKKTEVTQQGTFKESLPLKRNSDKL ARKKDPKKYGGFSDPVTAYSVLVVAKVEKG KSKKLKSVKELLGI TIMERS SFEKNPDLFLAEKYGKVEKVDL I I KLPKYSLEFENGRKLMASAGEL QKGNELAPLSYVLASYKELKSGPEDNEQKQLFVEQHKHYLDEI IEIQISEFSKRVILADANL KVLASAYNHKRDPLFIREQAEI IHFLTTLNLGAPAFAKYFDTTIDRKYSTKEVLATLHIQS ITGLYPETRTSLQGGLD (SEQ ID NO:37)

166
DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDFQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKRQTF
DDIPQHIGLHELHAILHRQDEYDFFQVFLGFVGLRNSRFWMTKSEITIPWP
NFEVVDKGAHASQFIERMTNDKPNLKEPLKHSLLYETYFVNTLETVKYEVMRKAFLPASSEQ
KKAIVDDILFKTNRKVQLKEDYFKKIECDFSVEI SGEVDRFNASGTYDHDLKI IDDKFDLDEEEM
EDIILVDITLFDEREMIELERTYHLAIDKVFQKLFQVMDLGKQNYDEELV
KVMGRKPHKENIEMARENQTQKQKNSRMRKIEEIGEIKLGOSILKEHPVENTQLONEKLYLYL
VQGRMDYVQEDLRSDLHVDHYVQPQFDKLDSDIDNKNLVRSDKRGSDNVSEPVEEKKMYW
QMLNKLITQRGFDNLTKAEELGDSLEKDAGKFQKLQVLVETRQITKHVAIQLDSSMRTKNEYDKLRE
VKVTILKSVKLSDRFQDYKFYKREINNYHAADHAYLNAVGTALIKYKPELSEFYDGYKVDVVRK
MIASKSEQIEKIGKAYFFSINMMNFKTEITLANGEIRKRPLIENGETGEI WVDKGDVFATVRKVL
MPQVINVKKETVEQTFGFSEILFKNSDHLI ARKDWDPKKYGDFSPTAYSVLVAKVEKGGSKK
LKCLAEGTRIFDPVFVTTRHEIDVDGKPFHVAAADGDTLLAEPVSWFDQGTRHQVILRIAGAI
VWATFDKVLTEYGWRAAGELRGKDGTLSSDQFYMGGNSLALSDTALADPepyDTPFSE
SEASMMGLLNALDRELVHVIINWAKRVPGFDVTDLDQAHELARWNELEIMILGWMMEHPGFKLFA
PNLLDDNQKCGECVEIDGKNSDFQKI ARKDWDPKKYGDFSPTAYSVLVAKVEKGGSKK
HIHRLDKITDLHLMKALGGTQLQQHQRLAILLLILSHHRMSKMGHELWMSYKVPNVLDDL
EMDLAHRLAGSSRGVRQAFAADLLLFLHDMLAEGLRYSVIREVIPTRARTFDLVEELHHTAVE
GVVNVHEKVEGLIIFQEDIFLEAKGYVEKKDL IIKLKYSFLEENGKRMLASAGEL
QKGNELALPSYKFVYFLASHYEKLGNASPDNEOKQFLVFEQHKKYDEL EIIEISEFMRVLADANLD
KVLSSAYNHRDPIREQAENI IHLFTLNLGAPAAK FYDHT DFRKSTEKVDLATLHQSI GTGLY
ETRIDLSQGDLK (SEQ ID NO:40)

Cas9: Intein (37R3-2; in double underline) replacing Ser1274

MDKKS IGLDIGTNSGVWAVIDTEYKYPKSKFKVLGNTDHS IKKNLIGALLFDSGTAETAEKLRKA
RRYTRKRNKRYLQEFISPMHLEESFVEDKHERHPFIGNI VDEAYEHKRYTIPY
HRRKLVDSTDKADRLI YLALAHMIFRGHLIEGDINPNDSVDKLLQIPQQLYQLEENNIPAS
GVDAKAILSRLSKRRLNELI AQLFGEKKNLGFIANI ALSLGLTIFPKSNFDLAEACKLQSKYDY
DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDFQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKRQTF
DDIPQHIGLHELHAILHRQDEYDFFQVFLGFVGLRNSRFWMTKSEITIPWP
NFEVVDKGAHASQFIERMTNDKPNLKEPLKHSLLYETYFVNTLETVKYEVMRKAFLPASSEQ
KKAIVDDILFKTNRKVQLKEDYFKKIECDFSVEI SGEVDRFNASGTYDHDLKI IDDKFDLDEEEM
EDIILVDITLFDEREMIELERTYHLAIDKVFQKLFQVMDLGKQNYDEELV
KVMGRKPHKENIEMARENQTQKQKNSRMRKIEEIGEIKLGOSILKEHPVENTQLONEKLYLYL
VQGRMDYVQEDLRSDLHVDHYVQPQFDKLDSDIDNKNLVRSDKRGSDNVSEPVEEKKMYW
QMLNKLITQRGFDNLTKAEELGDSLEKDAGKFQKLQVLVETRQITKHVAIQLDSSMRTKNEYDKLRE
VKVTILKSVKLSDRFQDYKFYKREINNYHAADHAYLNAVGTALIKYKPELSEFYDGYKVDVVRK
MIASKSEQIEKIGKAYFFSINMMNFKTEITLANGEIRKRPLIENGETGEI WVDKGDVFATVRKVL
MPQVINVKKETVEQTFGFSEILFKNSDHLI ARKDWDPKKYGDFSPTAYSVLVAKVEKGGSKK
LKCLAEGTRIFDPVFVTTRHEIDVDGKPFHVAAADGDTLLAEPVSWFDQGTRHQVILRIAGAI
VWATFDKVLTEYGWRAAGELRGKDGTLSSDQFYMGGNSLALSDTALADPepyDTPFSE
SEASMMGLLNALDRELVHVIINWAKRVPGFDVTDLDQAHELARWNELEIMILGWMMEHPGFKLFA
PNLLDDNQKCGECVEIDGKNSDFQKI ARKDWDPKKYGDFSPTAYSVLVAKVEKGGSKK
HIHRLDKITDLHLMKALGGTQLQQHQRLAILLLILSHHRMSKMGHELWMSYKVPNVLDDL
EMDLAHRLAGSSRGVRQAFAADLLLFLHDMLAEGLRYSVIREVIPTRARTFDLVEELHHTAVE
GVVNVHEKVEGLIIFQEDIFLEAKGYVEKKDL IIKLKYSFLEENGKRMLASAGEL
QKGNELALPSYKFVYFLASHYEKLGNASPDNEOKQFLVFEQHKKYDEL EIIEISEFMRVLADANLD
KVLSSAYNHRDPIREQAENI IHLFTLNLGAPAAK FYDHT DFRKSTEKVDLATLHQSI GTGLY
ETRIDLSQGDLK (SEQ ID NO:40)

DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDFQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKRQTF
DDIPQHIGLHELHAILHRQDEYDFFQVFLGFVGLRNSRFWMTKSEITIPWP
NFEVVDKGAHASQFIERMTNDKPNLKEPLKHSLLYETYFVNTLETVKYEVMRKAFLPASSEQ
KKAIVDDILFKTNRKVQLKEDYFKKIECDFSVEI SGEVDRFNASGTYDHDLKI IDDKFDLDEEEM
EDIILVDITLFDEREMIELERTYHLAIDKVFQKLFQVMDLGKQNYDEELV
KVMGRKPHKENIEMARENQTQKQKNSRMRKIEEIGEIKLGOSILKEHPVENTQLONEKLYLYL
VQGRMDYVQEDLRSDLHVDHYVQPQFDKLDSDIDNKNLVRSDKRGSDNVSEPVEEKKMYW
QMLNKLITQRGFDNLTKAEELGDSLEKDAGKFQKLQVLVETRQITKHVAIQLDSSMRTKNEYDKLRE
VKVTILKSVKLSDRFQDYKFYKREINNYHAADHAYLNAVGTALIKYKPELSEFYDGYKVDVVRK
MIASKSEQIEKIGKAYFFSINMMNFKTEITLANGEIRKRPLIENGETGEI WVDKGDVFATVRKVL
MPQVINVKKETVEQTFGFSEILFKNSDHLI ARKDWDPKKYGDFSPTAYSVLVAKVEKGGSKK
LKCLAEGTRIFDPVFVTTRHEIDVDGKPFHVAAADGDTLLAEPVSWFDQGTRHQVILRIAGAI
VWATFDKVLTEYGWRAAGELRGKDGTLSSDQFYMGGNSLALSDTALADPepyDTPFSE
SEASMMGLLNALDRELVHVIINWAKRVPGFDVTDLDQAHELARWNELEIMILGWMMEHPGFKLFA
PNLLDDNQKCGECVEIDGKNSDFQKI ARKDWDPKKYGDFSPTAYSVLVAKVEKGGSKK
HIHRLDKITDLHLMKALGGTQLQQHQRLAILLLILSHHRMSKMGHELWMSYKVPNVLDDL
EMDLAHRLAGSSRGVRQAFAADLLLFLHDMLAEGLRYSVIREVIPTRARTFDLVEELHHTAVE
GVVNVHEKVEGLIIFQEDIFLEAKGYVEKKDL IIKLKYSFLEENGKRMLASAGEL
QKGNELALPSYKFVYFLASHYEKLGNASPDNEOKQFLVFEQHKKYDEL EIIEISEFMRVLADANLD
KVLSSAYNHRDPIREQAENI IHLFTLNLGAPAAK FYDHT DFRKSTEKVDLATLHQSI GTGLY
ETRIDLSQGDLK (SEQ ID NO:40)
In some embodiments, the intein inserted into the Cas9 protein is ligand-dependent. In some embodiments, the ligand-dependent inteins comprise a modified ligand-binding domain of the estrogen receptor protein, embedded into a modified RecA intein from *M. tuberculosis*. In some embodiments, the ligand-binding domain is derived from an estrogen receptor protein, for example, from the human estrogen receptor. The sequence of the human estrogen receptor and the location of the ligand-binding domain within the human estrogen receptor are known to those of skill in the art. Non-limiting, exemplary sequences of the human estrogen receptor can be retrieved from RefSeq database entries NP_000116 (isoform 1); NP_001116212 (isoform 2); NP_001116213 (isoform 3); and NP_001116214 (isoform 4) from the National Center for Biotechnology Information (NCBI, www.ncbi.nlm.nih.gov). In some embodiments, the ligand-binding domain of a ligand-dependent intein provided herein comprises or is derived from a sequence comprising amino acid residues 304-551 of the human estrogen receptor.


[0083]  Additional exemplary inteins, ligand-binding domains, and ligands suitable for use in the Cas9 proteins provided herein are described in International Patent Application, PCT/US20 12/028435, entitled "Small Molecule-Dependent Inteins and Uses Thereof," filed March 9, 2012, and published as WO 2012/125445 on September 20, 2012, the entire contents of which are incorporated herein by reference. Additional suitable inteins, ligand-binding domains, and ligands will be apparent to the skilled artisan based on this disclosure.

[0084]  The ligand-dependent inteins provided herein are inactive (or only minimally active) in the absence of the appropriate ligand, but can be induced to be active, and, thus, to self-excite, by contacting them with a ligand that binds the ligand-binding domain of the human estrogen receptor. Small molecule ligands binding the ligand-binding domain of the estrogen receptor (e.g., the human estrogen receptor), and thus useful to induce the activity of the ligand-dependent inteins described herein, are known to those of skill in the art. In some embodiments, the ligand used to induce the activity of the ligand-dependent inteins described herein specifically binds to the ligand-binding domain of the estrogen receptor. In some embodiments, the ligand binds the ligand-binding domain of a ligand-dependent intein
provided herein with high affinity, for example, with an affinity of at least about $10^{-10}$ M, at least about $10^{-9}$ M, at least about $10^{-8}$ M, at least about $10^{-7}$ M, or at least about $10^{-6}$ M. Examples of appropriate estrogen receptor-binding ligands that are useful to induce the activity of the ligand-dependent inteins provided herein, for example, the ligand-dependent inteins provided in SEQ ID NOs 3-8, include, but are not limited to, 17P-estradiol, 17a-ethyl estradiol, tamoxifen and tamoxifen analogs (e.g., 4-hydroxytamoxifen (4-HT, 4-OHT), 3-hydroxytamoxifen (droloxifene)), tamoxifen metabolites (e.g., hydroxytamoxifen, endoxifen), raloxifene, toremifene, ICI-182, and ICI-780. Other useful ligands will be apparent to those of skill in the art, and the invention is not limited in this respect.

[0085] In some embodiments, any of the Cas9 proteins comprising inteins (e.g., SEQ ID NOs:27-41) can be modified so as to generate a Cas9 nickase comprising an intein (e.g., by making one of a D10A or H840A mutation relative to the Cas9 sequence lacking an intein), or to generate a dCas9 protein comprising an intein (e.g., by making both D10A and H840A mutations relative to the Cas9 sequence lacking an intein). In some embodiments, any of the Cas9 proteins comprising inteins (e.g., SEQ ID NOs:27-41) have additional features, for example, one or more linker sequences, localization sequences, such as nuclear localization sequences (NLS; e.g., MAPKKKRKVGIHRGVP (SEQ ID NO:42)); cytoplasmic localization sequences; export sequences, such as nuclear export sequences; or other localization sequences, as well as sequence tags that are useful for solubilization, purification, or detection of the fusion proteins. Suitable localization signal sequences and sequences of protein tags are provided herein and are known in the art, and include, but are not limited to, biotin carboxylase carrier protein (BCCP) tags, myc-tags, calmodulin-tags, FLAG-tags (e.g., 3xFLAG TAG: MDYKDHGDYKDHIDYKDDDDK (SEQ ID NO:43)), hemagglutinin (HA) tags, polyhistidine tags, also referred to as histidine tags or His-tags, maltose binding protein (MBP)-tags, nus-tags, glutathione-S-transferase (GST) tags, green fluorescent protein (GFP) tags, thioredoxin-tags, S-tags, Softags (e.g., Softag 1, Softag 3), strep-tags, biotin ligase tags, F1AsH tags, V5 tags, and SBP-tags. Additional suitable sequences will be apparent to those of skill in the art.

[0086] In some embodiments, ligand-dependent site-specific enzymes (e.g., fusion proteins) are provided which comprise a Cas9 variant (e.g., dCas9), a ligand-dependent intein, and one or more other polypeptide domains having a particular enzymatic activity. In some embodiments, the fusion protein comprises a nuclease inactivated Cas9 domain (e.g., dCas9), wherein the dCas9 domain comprises an intein sequence inserted in place of or in
addition to any amino acid in dCas9. In some embodiments the inserted intein sequence replaces Cys80, Alal27, Thr146, Ser219, Thr333, Thr519, Cys574, Thr622, Ser701, Ala728, Thr995, Serl006, Serl54, Serl59, or Serl274 of dCas9 (SEQ ID NO:5). In some embodiments, the inserted intein sequence replaces Alal27, Thr146, Ser219, Thr519, or Cys574 of dCas9 (SEQ ID NO:5). In some embodiments, the intein is inserted into another domain of the fusion protein (i.e., not in the Cas9 domain, e.g., not in the dCas9 domain), such as the domain having a particular enzymatic activity. In some embodiments, the domain having a particular enzymatic activity is a nuclease domain (e.g., FokI), a recombinase catalytic domain (e.g., Hin, Gin, or Tn3 recombinase domains), a nucleic acid-editing domain (e.g., a deaminase domain), a transcriptional activator domain (e.g., VP64, p65), a transcriptional repressor domain (e.g., KRAB, SID), or an epigenetic modifier (e.g., LSD1 histone demethylase, TET1 hydroxylase). The intein that is inserted into the fusion protein can be any ligand-dependent intein, e.g., those described herein. For example, in some embodiments, the intein that is inserted into a Cas9 protein comprises in part or in whole, a sequence that is at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% percent identical to any one of SEQ ID NO:7-14.

In some embodiments, the general architecture of exemplary fusion proteins provided herein comprises the structure:

\[ \text{NH}_2-[\text{enzymatic domain}]-[\text{dCas9}]-[\text{COOH}] \]  
\[ \text{NH}_2-[[\text{dCas9}]-[\text{COOH}] \]

wherein \( \text{NH}_2 \) is the N-terminus of the fusion protein, COOH is the C-terminus of the fusion protein, dCas9 comprises an intein as provided herein, and the enzymatic domain comprises a nuclease domain (e.g., FokI), a recombinase catalytic domain (e.g., Hin, Gin, or Tn3 recombinase domains), a nucleic acid-editing domain (e.g., a deaminase domain), a transcriptional activator domain (e.g., VP64, p65), a transcriptional repressor domain (e.g., KRAB, SID), or an epigenetic modifier (e.g., LSD1 histone demethylase, TET1 hydroxylase). In some embodiments, the intein is comprised in a domain other than dCas9 (e.g., in an enzymatic domain), or is located between two domains.

Additional features may be present, for example, one or more linker sequences between certain domains. Other exemplary features that may be present are localization sequences, such as nuclear localization sequences (NLS; e.g., MAPKKKRKVGIGHRGVP (SEQ ID NO:42)); cytoplasmic localization sequences; export sequences, such as nuclear export sequences; or other localization sequences, as well as sequence tags that are useful for solubilization, purification, or detection of the fusion proteins. Suitable localization signal
sequences and sequences of protein tags are provided herein and are known in the art, and include, but are not limited to, biotin carboxylase carrier protein (BCCP) tags, myc-tags, calmodulin-tags, FLAG-tags (e.g., 3xFLAG TAG: MDYKHDGDYKDHDIDYKDDDDK (SEQ ID NO:43)), hemagglutinin (HA) tags, polyhistidine tags, also referred to as histidine tags or His-tags, maltose binding protein (MBP)-tags, nus-tags, glutathione-S-transferase (GST) tags, green fluorescent protein (GFP) tags, thioredoxin-tags, S-tags, Softags (e.g., Softag 1, Softag 3), strep-tags, biotin ligase tags, FlAsH tags, V5 tags, and SBP-tags. Additional suitable sequences will be apparent to those of skill in the art.

In some embodiments, the enzymatic domain comprises a nuclease or a catalytic domain thereof. For example, in some embodiments, the general architecture of exemplary ligand-dependent dCas9 fusion proteins with a nuclease domain comprises the structure:

\[
\text{[NH}_2\text{-][NLS]-[dCas9]-[nuclease]-[COOH]}, \\
\text{[NH}_2\text{-][NLS]-[nuclease]-[dCas9]-[COOH]}, \\
\text{[NH}_2\text{-}[dCas9]-[nuclease]-[COOH]}, \text{or} \\
\text{[NH}_2\text{-}[nuclease]-[dCas9]-[COOH]};
\]

wherein NLS is a nuclear localization signal, dCas9 comprises an intein as provided herein, NH\textsubscript{2} is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein. In some embodiments, a linker is inserted between the dCas9 and the nuclease domain. In some embodiments, a linker is inserted between the NLS and the nuclease and/or dCas9 domain. In some embodiments, the NLS is located C-terminal of the nuclease and/or the dCas9 domain. In some embodiments, the NLS is located between the nuclease and the dCas9 domain. Additional features, such as sequence tags, may also be present. In some aspects, the nuclease domain is a nuclease requiring dimerization (e.g., the coming together of two monomers of the nuclease) in order to cleave a target nucleic acid (e.g., DNA). In some embodiments, the nuclease domain is a monomer of the FokI DNA cleavage domain. The FokI DNA cleavage domain is known, and in some aspects corresponds to amino acids 388-583 of FokI (NCBI accession number J04623). In some embodiments, the FokI DNA cleavage domain corresponds to amino acids 300-583, 320-583, 340-583, or 360-583 of FokI. See also Wah et al, "Structure of FokI has implications for DNA cleavage" Proc. Natl. Acad. Sci. USA. 1998; 1;95(18): 10564-9; Li et al, "TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain" Nucleic Acids Res. 2011; 39(l):359-72; Kim et al, "Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain" Proc. Natl Acad. Sci. USA. 1996; 93:1156-1160; the entire contents of each are
herein incorporated by reference). In some embodiments, the FokI DNA cleavage domain corresponds to, or comprises in part or whole, the amino acid sequence set forth as SEQ ID NO:6. In some embodiments, the FokI DNA cleavage domain is a variant of FokI (e.g., a variant of SEQ ID NO:6), as described herein. Other exemplary compositions and methods of using dCas9-nuclease fusion proteins can be found in U.S. patent application U.S.S.N 14/320,498; titled "Cas9-FokI fusion Proteins and Uses Thereof," filed June 30, 2014; the entire contents of which are incorporated herein by reference.

[0090] In some embodiments, the enzymatic domain comprises a recombinase or other catalytic domain thereof. For example, in some embodiments, the general architecture of exemplary ligand-dependent dCas9 fusion proteins with a recombinase domain comprises the structure:

\[
[NH_2][NLS][dCas9][recombinase][COOH],
\]

\[
[NH_2][NLS][recombinase][dCas9][COOH],
\]

\[
[NH_2][dCas9][recombinase][COOH],
\]

\[
[NH_2][recombinase][dCas9][COOH];
\]

wherein NLS is a nuclear localization signal, dCas9 comprises an intein as provided herein, \(NH_2\) is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein. In some embodiments, a linker is inserted between the dCas9 and the recombinase domain. In some embodiments, a linker is inserted between the NLS and the recombinase and/or dCas9 domain. In some embodiments, the NLS is located C-terminal of the recombinase domain and/or the dCas9 domain. In some embodiments, the NLS is located between the recombinase domain and the dCas9 domain. Additional features, such as sequence tags, may also be present. By "catalytic domain of a recombinase," it is meant that a fusion protein includes a domain comprising an amino acid sequence of (e.g., derived from) a recombinase, such that the domain is sufficient to induce recombination when contacted with a target nucleic acid (either alone or with additional factors including other recombinase catalytic domains which may or may not form part of the fusion protein). In some embodiments, a catalytic domain of a recombinase does not include the DNA binding domain of the recombinase. In some embodiments, the catalytic domain of a recombinase includes part or all of a recombinase, e.g., the catalytic domain may include a recombinase domain and a DNA binding domain, or parts thereof, or the catalytic domain may include a recombinase domain and a DNA binding domain that is mutated or truncated to abolish DNA binding activity. Recombinases and catalytic domains of recombinases are known to those of skill in the art, and include, for example, those described herein. In some embodiments, the
catalytic domain is derived from any recombinase. In some embodiments, the recombinase catalytic domain is a catalytic domain of a Tn3 resolvase, a Hin recombinase, or a Gin recombinase. In some embodiments, the catalytic domain comprises a Tn3 resolvase (e.g., Stark Tn3 recombinase) that is encoded by a nucleotide sequence comprising, in part or in whole, SEQ ID NO:44, as provided below. In some embodiments, a Tn3 catalytic domain is encoded by a variant of SEQ ID NO:44. In some embodiments, a Tn3 catalytic domain is encoded by a polynucleotide (or a variant thereof) that encodes the polypeptide corresponding to SEQ ID NO:45. In some embodiments, the catalytic domain comprises a Hin recombinase that is encoded by a nucleotide sequence comprising, in part or in whole, SEQ ID NO:46, as provided below. In some embodiments, a Hin catalytic domain is encoded by a variant of SEQ ID NO:46. In some embodiments, a Hin catalytic domain is encoded by a polynucleotide (or a variant thereof) that encodes the polypeptide corresponding to SEQ ID NO:47. In some embodiments, the catalytic domain comprises a Gin recombinase (e.g., Gin beta recombinase) that is encoded by a nucleotide sequence comprising, in part or in whole, SEQ ID NO:48, as provided below. In some embodiments, a Gin catalytic domain is encoded by a variant of SEQ ID NO:48. In some embodiments, a Gin catalytic domain is encoded by a polynucleotide (or a variant thereof) that encodes the polypeptide corresponding to SEQ ID NO:49. Other exemplary compositions and methods of using dCas9-recombinase fusion proteins can be found in U.S. patent application U.S.S.N 14/320,467; titled "Cas9 Variants and Uses Thereof," filed June 30, 2014; the entire contents of which are incorporated herein by reference.

Stark Tn3 recombinase (nucleotide: SEQ ID NO:44; amino acid: SEQ ID NO:45):
ATGGCCCTGTGTGGCTACGCACGGGGTGCACATTTCCATTCTCTGACCGATTCCATCGATACGTCTAGCGCC
AAGTACGTGAATAAGGGCGATACTCTGGTTGTGTGGAAGTTGGATCGCTTGGGTAGATCAGTGAAGAATCTCGTA

Hin Recombinase (nucleotide: SEQ ID NO:46; amino acid: SEQ ID NO:47):
ATGGCAACCATTTGGCTCACAATAGGGTGTCTACCATCGACGCACGCAACGCTCTGACACATCC
GCCAAGCCTGCGATCTCTTCAGGATGAAGTACGTCTGGCGCAAGATCGCAGGCAAGCTCTCCTAGGGCTCTG
AAGTACGTTGAATAGGGCCGATACTGGTTTGGTTGGATGGATCTGCTGATGGAAGAATCTGGTA
GCCCTGATAAGCGAGCTGCAAGGGGTGGTCACATTTCCATTCTCCTGACCATCTTACGTAGCTAGCGCC
ATGGGCCGATTCTTCTTTTACGTCATGTCCGCCCTCGCTGAAATGGAGCGCGAACTTATTGTTGAACGGACTTTG
GCTGGACTGGCAGCGGCTAGAGCACAGGGCCGACTTGGA (SEQ ID NO:46)

MATIGYIRVST DQ01 DLQRNALTSANCDRIFEDRI SGKIANPGLKRALKNVKGDTLVWKLDRLGVSVPNLY
AL1 SELHERGAHFHTLTD S DTS SMARGFFYYVMSALAEMEREL IVERTLAGLAARAOGRILG (SEQ ID
NO:47)

Gin beta recombinase (nucleotide: SEQ ID NO:48; amino acid: SEQ ID NO:49):
ATGCTCATTGGCTATGTAAGGGTCAGCACCAATGACCAAAACACAGACTTGCAACGCAATGCTTTGGTTTGCGCC
GGATGTGAACAGATATTTGAAGATAAACTGAGCGGCACTCGGACAGACAGACCTGGGCTTAAGAGAGCACTGAAA
AGACTGCAGAAGGGGACACCCCTGGTCGTCTGGAAACTGGATCGCCTCGGACGCAGCATGAAACATCTGATTAGC
CTGGTTGGTAGCTTAGGAGGAGAGAAATCAACTTCAGAAGGGGGACACCCTGGTCGTCTGGAAACTGGATCGCCTCGG
ACGCAGCATGAAACATCTGATTAGCCTGGTTGGTAGCTTAGGAGGAGAGAAATCAACTTCAGAAGGGGGACACCCTGG
TCGTCTGGAAACTGGATCGCCTCGGACGCAGCATGAAACATCTGATTAGCCTGGTTGGTAGCTTAGGAGGAGAGAA
ATCACTTCAGAAGGGGGACACCCTGGTCGTCTGGAAACTGGATCGCCTCGGACGCAGCATGAAACATCTGATTAGCCTG
GATCGCTGCTGCCGGAAACAAAGGCAGACGGTTCGGCAAGACCGCCGAAGAGCGGC (SEQ ID NO:48)

ML IGYVRVSTNDQNTDQLVRNALVCAGCEQIFEDKLSGTRTDRPLKLRLRKLQKGDTLVVWKLDRLGRSKHFL 1 S
LVGLERERGINFSLTD S DTS SPMGRFFYYVMGALAEEREL IERTMAGIAARNGRFRPGPKSG (SEQ ID
NO:49)

[0091] In some embodiments, the enzymatic domain comprises a deaminase or a
catalytic domain thereof. For example, in some embodiments, the general architecture of
exemplary dCas9 fusion proteins with a deaminase enzyme or domain comprises the
structure:

\[ \text{[NH}_2\text{-NLS-}[Cas9]-[deaminase]-[COOH]}, \]
\[ \text{[NH}_2\text{-NLS-}[deaminase]-[Cas9]-[COOH]}, \]
\[ \text{[NH}_2\text{-}[Cas9]-[deaminase]-[COOH]}, \]
\[ \text{[NH}_2\text{-deaminase}-[Cas9]-[COOH]}, \]

wherein NLS is a nuclear localization signal, dCas9 comprises an intein as provided herein,
NH2 is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion
protein. In some embodiments, a linker is inserted between the dCas9 and the deaminase
domain. In some embodiments, a linker is inserted between the NLS and the deaminase
and/or dCas9 domain. In some embodiments, the NLS is located C-terminal of the
deaminase and/or the dCas9 domain. In some embodiments, the NLS is located between the
deaminase domain and the dCas9 domain. Additional features, such as sequence tags, may
also be present. One exemplary suitable type of nucleic acid-editing enzymes and domains
are cytosine deaminases, for example, of the apolipoprotein B mRNA-editing complex
(APOBEC) family of cytosine deaminase enzymes, including activation-induced cytidine
deaminase (AID) and apolipoprotein B editing complex 3 (APOBEC3) enzyme. Another
exemplary suitable type of nucleic acid-editing enzyme and domain thereof suitable for use in
the present invention include adenosine deaminases. For example, an ADAT family
adenosine deaminase can be fused to a dCas9 domain comprising an intein. Some exemplary suitable nucleic-acid editing enzymes and domains, e.g., deaminases and deaminase domains, that can be fused to dCas9 domains comprising inteins according to aspects of this disclosure are provided below. It will be understood that, in some embodiments, the active domain of the respective sequence can be used, e.g., the domain without a localizing signal (nuclear localizing signal, without nuclear export signal, cytoplasmic localizing signal). Other exemplary compositions and methods of using dCas9-nuclease fusion proteins can be found in U.S. patent application U.S.S.N 14/325,815; titled "Fusions of Cas9 Domains and Nucleic Acid-Editing Domains," filed July 8, 2014; the entire contents of which are incorporated herein by reference.

Human AID:
MDSLLMNRRKFLYQFKNVRAKGRHETLYC YVVKR RDSATSFSDLFGYLRNKNGCHVELLFLRYI SDWLDPGRC YRVTVTFSWSPCYDCAHVAETFRLGNP NLSLRI FTARLYFECRKAEPGLRLHRAGVQIAIMTFKDYFCWNT FVENHERTFKAWEGLHENSVRLRQLRRILLP LYEVD LRDARFL TLGL (SEQ ID NO:50)
(underline: nuclear localization signal; double underline: nuclear export signal)

Mouse AID:
MPS LLNQKFKFLYHFKNVRAKGRHETLYC YVVKR RDSATSCSDLFGHLRNKSGCHVELLFLRY ISDWLDPGRC YRVTVTFSWSPCYDCAHVEFLRNPNLSLRI FTARLYFECRKAEPGLRLHRAGVQIGMTFKDYFCWNT FVENHREKTFKAWEGLHENSVRLRQLRRILLP LYEVDLRLRAFLMLGLF (SEQ ID NO:51)
(underline: nuclear localization signal; double underline: nuclear export signal)

Dog AID:
MPS LLNQKFRKFLYHFKNVRAKGRHETLYC YVVKR RDSATSFSDLFGHLRNKSGCHVELLFLRY ISDWLDPGRC YRVTVTFSWSPCYDCAHVFADFLGNP NLSLRI FTARLYFECRKAEPGLRLHRAGVQIAIMTFKDYFCWNT FVENHREKTFKAWEGLHENSVRLRQLRRILLP LYEVDLRLRAFRITLGL (SEQ ID NO:52)
(underline: nuclear localization signal; double underline: nuclear export signal)

Bovine AID:
MPSLLKQKFQFLYQFKNVRAKGRHETLYC YVVKR RPSSTSFSDLFGHLRNKAGCHVELLFLRYI SPWPPLPGRC YRVTVTFSWSPCYFCAHAPFLEYPNLSLRI FTARLYFCRKAEPGLRLHRAGVQIAIMTFKPYFCWNT TFVENHRTFKAWEGLHENSVRLRQLRRILLP LYEVDLRLPAFLTLVLGL (SEQ ID NO:53)
(underline: nuclear localization signal; double underline: nuclear export signal)

Mouse APOBEC-3:
MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYA KGRKPTFLCYYEVTTRCPSPVSLHKGVFKNPN I HAEIC FLYWFWHPKVLVSLPREEFKITWMSPSFC EC AEQIVFRLATHTNLSPI FSSRLYNQPE TQQNLCLLVQEG AQVAAMPMYF KCCWKFVPGGRFRWPWKRTLNNFRYQPSKQLQEI LRPCYIVPSSSSTLSNICTKGTLPE TR FCVEGRRMPLSSEEFYSQFYQVRKHLCYYHRMKP L CYQLQFNGQAFLKGC LSEKGKQAELFLPKIRSM
ELSQVT I TCYLTWSPCPNC AWQLAAFKRDRPDL I LHIYTSRLYFHWKRPFQKGLCSLWQSGI LVDVMDLPQFTDC
WTNFVNPKRPFPWPKGLE I ISSRTQRRLRRIKESWQLDVLNDQFLNGQGPLMS (SEQ ID NO:54)
(underline: nucleic acid editing domain)

Rat APOBEC-3:
MGPFCLGCSHRKCYSP IRNL I SQETFKFHFKNRLRYAI DRKDTFLCYEVTRKDCDSPVSLIHGVFKNNDNII HAEEIC
FLYWYFHDKVLKVSPPREEFKI TWYMSWSCFCAGAEQVRFLATHIFNLSDDF SRLYNIREDPENQNLCLRLVQEG
AQVAAMDLYEFFKCKWKFVNDNGRRFPWPKLII TFNRYQDSKLQEI I LRPCYIPVPS S S STLSNICTKCLPETR
FCVERRVHLLLSEEFEYSFQYNQVRKLYCYYGKKPLYCQLQEQFNGQAPLKGCLLSEKGGKHAEIEILFKKIRSM
ELSQVI I TCYLTWSPCPNC AWQLAIFKDRPDL I LHIYTSRLYFHWKRPFQKGLCSLWQSGI LVDVMDLPQFTDC
WTNFVNPKRPFPWPKGLE I ISSRTQRRLRRIKESWQLDVLNDQFLNGQGPLMS (SEQ ID NO:55)
(underline: nucleic acid editing domain)

Rhesus macaque APOBEC-3G:
MVEPMDPRTFVSFNRRNP I LSLGLNTWVLCEEVKTISGPPLDAAIFQGKYSSAKYHPg * RFLRWFHWRQPLAH
DQEUKEYTVWYVSIVSPCirCANSVATFALAKDPKVTTL IFVARLILYFVKPDYQQALRI LCQKRRGPHTAMKIMNYNEF
QDCWKNFVDRGKPKFPRNLPKHYTLLQATLGERLRLHMDGPTFSNFKPWSQYQHETLYCYPVERLHNDTW
VPLNQHRGFLRNAPNIGFPKRGHAELCFDLIPFICILDGecerIYrFCFISVSPC SFCAQEMAKF I SNEHVS1CL
IFAAIRI YDDFOGRYQEGLRALHRDGAIAMNYESFEYCVWDFTQFRPFQPDGLDEHSQALSGRLRAI
(SEQ ID NO:56)
(bold italic: nucleic acid editing domain; underline: cytoplasmic localization signal)

Chimpanzee APOBEC-3G:
MKPHFRNPVEMQYDQTSDFNYRNRP I LSHTNTWVLCEVEKTKGSPPLDAAIFRGQQYSSAKYHPEMRFHWF
KWRKLHRDOEYTLYWSWPSCTKCRDVTAVLFADPVKVTTL IFVARLILYFVKPDYQQALRI LCQERGGPHATMKIMNYNEF
QDCWKNFVDRGKPKFPRNLPKHYTLLQATLGERLRLHMDGPTFSNFKPWSQYQHETLYCYPVERLHNDTW
VPLNQHRGFLRNAPNIGFPKRGHAELCFDLIPFICILDGecerIYrFCFISVSPC SFCAQEMAKF I SNNKHVS1CL
IFAAIRI YDDFOGRYQEGLRALHRDGAIAMNYESFEYCVWDFTQFRPFQPDGLDEHSQALSGRLRAI
(SEQ ID NO:57)
(underline: nucleic acid editing domain; double underline: cytoplasmic localization signal)

Green monkey APOBEC-3G:
MNPQIRNVEQEMPEDYVYFVINRNRP I LSGRNTRWVLCVEVKTKDPSGPPLDAAIFRGQQYSSAKYHPEA KDHPEMKFLHWFR
KWRKLRDOEYTLYWSWPSCTKCRDVTAVLFADPVKVTTL IFVARLILYFVKPDYQQALRI LCQERGGPHATMKIMNYNEF
QDCWKNFVDRGKPKFPRNLPKHYTLLQATLGERLRLHMDGPTFSNFKPWSQYQHETLYCYPVERLHNDTW
VPLNQHRGFLRNAPNIGFPKRGHAELCFDLIPFICILDGecerIYrFCFISVSPC SFCAQEMAKF I SNNKHVS1CL
IFAAIRI YDDFOGRYQEGLRALHRDGAIAMNYESFEYCVWDFTQFRPFQPDGLDEHSQALSGRLRAI
(SEQ ID NO:58)
(underline: nucleic acid editing domain; double underline: cytoplasmic localization signal)
Human APOBEC-3G:
MKPHFRNTVERMYRDTFSYNFYNRPILSRNNTVWLCEYEVKTGSPSRPPLDAKIFRGQVYSEQ_E
LKYHPEMRFFFHDFWSKRNKRHLRQEPQYETVWYIWNPCTKCTRDATMAFLAEDPKVLITIFVARLYFWPDYQEAALARLSQCKRDKPRATMK
IMNYDEEQHXSWFKQYQRELFEWPNNLKYIILLHIMHGLIRHSMIDPFTFTFNNNPWRGHRHETYLYCEVE
RMHNHTTWLNLNQRPFRGCNQAPHKGFLEGRAWELFVILPKLDDQMYRVTUTCSTWSPCFCSCAQEMAKFIS
KNKHVSLCFITARIYDDQQRCQCELRALTAEAGAKISIMTYSEFKHCWDTFVHGQCPFPWDGLDESHQDLGSR
RAiLQNQEN (SEQ ID NO:59)
(underline: nucleic acid editing domain; double underline: cytoplasmic localization signal)

Human APOBEC-3F:
MKPHFRNTVERMYRDTFSYNFYNRPILSRNNTVWLCEYEVKTGSPSRPPLDAKIFRGQVYSEQ_E
LKYHPEMRFFHWFS
KWRKLHRDQEYEVTWYI... (SEQ ID NO:60)
(underline: nucleic acid editing domain)

Human APOBEC-3B:
MNQIRNPMEYMYRTDFYDNFENEPILYGRSYTWLCEYEVKIRGRSNLWDTGVFRGQVYFPQ_HAEACFLSWFC
GQNLPAFYKFQITFWVSVTFPCDCVAKLAEFLAEHPNVTLTI_SAARLYYYWERDYYYRALCRLSQAARVTIMDY
EEFYCWENFYVNEQGQENMFWQETYRFLHTKELRRNPEAMYPHIFYHFKNKLKAYGRENESLWCLFTMEVV
KHSPVSWKGRVFRNQVDPETHC_HAERCFLSWFCDDILSPNTYEVTWYTSNPCPECAGEVAEFRHALSHSVNLT
IFTARLYFWTDYQEGLRSLSEQASGEVIMGKYDFKYCWFENFYNDDEFPKPWKGLYNFLDSDKLEILQEILE
(SEQ ID NO:61)
(underline: nucleic acid editing domain)

Human APOBEC-3C:
MNQIRNPKMAYMGTFNYQFKNLWEANDRNETWCLFTVEGIKRRSVWSTGVFRNQVDSETH_HAERCFLSWFC
CDDILSPNTKYQITWYTSSPCPDCAEVEAFALLRNASNVTILITFTARLYFYQPCYQEGLRSLSEQGAVEIMDY
EDFKYCWFENFYNDNEFPKPWKGLKTNFRLKRRLLRESLQ (SEQ ID NO:62)
(underline: nucleic acid editing domain)

Human APOBEC-3A:
MEASPASGPRHLMDFPHTSNFNNNGIRKHYTLYCVEVERLDNGTSVMDQHRGFLHNQAKNLLCGFYGR_HAEELRF
LDLVPSSLQDPAIQYRTWFIWGSPCFSWGCAGEVRALQENTHVRRLFAARIYDDPLYKEALQMLRADAQGV
SIMTYSEFKHCWDTFVHGQCPFPWDGLDESHQALSGLRLRAiLQNQEN (SEQ ID NO:63)
(underline: nucleic acid editing domain)

Human APOBEC-3H:
MALLTATFLKFNNKKRLLRPYYPKALLCCYQLTPQNGSTPRGNYFENKKC_HAEICFINEIKSMGLDTEQCYQ
VTCYLTWSPCSSCAWEMLVDFTIKAHDLNLGGIFASRIYLYWHKCPQQRGLLRCGQSVPEVEGMPFKADCWENVD
HEKPLSFNYKMLLEDKNSAIIRKRLRIKIPGVRAGQRYMIDLCDAEV (SEQ ID NO:64)
(underline: nucleic acid editing domain)
Human APOBEC-3D:
MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGPVLPKRQSNHRQEVYFRFEN
HAEMCFLSFWFCGNRLPANRFRQITWEVSNWNCPCLPVKVTFLAEHPNVITII SAAARLYYRDRDORWWRLRR
HKAGARV.KIDYEFDAYCWeNFCQNQFPFIWMFYYKFDNNAYSLHRTLKEILRNPKMEMAMYHPHYYFHKNLKKACG
RNESWLCTFVMETVHKSHASVRKGFVRQNVQDFPETHCHAERCFLSWFCDDILSPNTNYEVTWTSWPSCPQAEVE
AEFLARHSHNVNLTIIFARLCYWFDTDYQECGLSLSQEGAVKIMGYKDFVSCWKNFVYSDDEPFPKPWKGLQNFR
LLKRLREILQ (SEQ ID NO:65) (underline: nucleic acid editing domain)

Human APOBEC-1:
MTSEKGPSTGDPTLRRRIPFEFEDVFDYPRELRLKEACLLYIEIKWGMRSKWRSSGKNNTNHEVNF1IKFSTSRDF
HPMSMCSITWLFSWSPCWECSEQQLFERLSRPVGLTVIYVARLFWHMDOQNRGQLDLSVNSGTVQTIMRASEYY
HCWRFVNYPPGDEAHKQPYPPLWMLYALELCHIJLSPCLKI SRRWQNHTFFRLHLQNYQYQTIPPHILLA
TGLIHPSTVAWR (SEQ ID NO:66)

Mouse APOBEC-3:
MSSETGPVAVDPTLRRRIPHEFEVFFDVFDPRELRTCLLYEINWGRHSWWRHTSQNTSNHEVNFLEKFTTERY
FRPNTCRSTWFSLSPGCERSAITEFLSPHYTVLFYIYIARLYHTDQRNRQGLRLISSGTVQITM1EQEYC
YCWRFVNYPPSNEAYWPRYPLWMLVLYLELYCIGLSPCLKILRRKQPLTFFTITALQCSHYQRLPHILLWA
TGL (SEQ ID NO:67)

Rat APOBEC-1:
MSSETGPVAVDPTLRRRIPHEFEVFFDVFDPRELRTCLLYEINWGRHS IWRHSQNTNKHHEVNFIEKFTTERY
FCPNTCRSTWFSLSPGCERSAITEFLSPHYTVLFYIYIARLYHTDQRNRQGLRLISSGTVQITM1ESESG
YCWRFVNYPPSNEAYWPRYPLWMLVLYLELYCIGLSPCLKILRRKQPLTFFTITALQCSHYQRLPHILLWA
TGL (SEQ ID NO:68)

Human ADAT-2:
MEAKAPKGPAASGCASVSAAEETKMEEAMMAKALENTEVFPGCLMVYNNEVVGKRGNEVTNQKNATRAEMV
AIDQVLDWCRQSKSPSEVEFHTVLIVTVEPCICMCAALRLMKIPLVYGCQNERFGGCGSVLNIASADLPNGTR
PFQCIPGYRAEVEAMKLFTFYQKENPNAPSKVRKKEQKS (SEQ ID NO:69)

Mouse ADAT-2:
MEEKVESTTSFPPCVPVSVQETKEWMEEMAMRAKALENIEVFPGCLMVYNNEVVGKRGNEVTNQKNATRAEMV
AIDQVLDWCHQHQSPSTFHEFHTVLIVTVEPCICMCAALRLMKIPLVYGCQNERFGGCGSVLNIASADLPNGTR
PFQCIPGYRAEVEAELKLFTFYQKENPNAPSKVRKKEQKS (SEQ ID NO:70)

Mouse ADAT-1:
MWTADEIAQLCAYANWRLPKQGKPEPNREWTLLAALVVIKQASANQADIPEKEVQTKEVVE SMGTMGKCIQGSK
MREGSGLDINSHAEI1AIASSFQKPLLHLHQLLAAVLEDSIFVPGTQTQGLRPLDLSFVFTHCPGAS11PP
LEFEQEPCCCPVRWNASNPNPQETQENLEDSDKPRNCEDPASVAKMMRGTPARLSNCVAHHTQEGSEGFVKPVV
SSSDSLTKEEFANDIGASGWRVQTVYRTAGCVPFTGDLRFQAGYCHQVLLRVRKPGRTDTSMSCDKMAR
WWNLGCQGALLMHFEKPLYSALVIKPGCPEQAMARRALTGCREEFLVPRFGVQLEI0QSGLLFEQSGRCGV
HRRKGPSFRGFLPVCAGAI SWASVPGPQDLVTTANGFGQDTKEIGSPRASRI SKVETRFSQKLSSAASDEQP
DSiRVKIKLTDYQYKDAASYAQEGWALRRQFASwiRNPPDYHQFK (SEQ ID NO:71) (underline: nucleic acid editing domain)
Human ADAT-1:

MWTADE IAQLCYEHYGIRLPKKGKPEPNHEWTLLAAVVKIQSPADKACDTPDKPVQVTKEVV SMGTGT KC IGQSK MRKNGDI LNDSHAEVIPARRSFQRYLLHQQLAATL KEDS IFVPTQKGVWKLRLRD IFVFFS SHTCPG DAS1 IPM LEFEDQCPFRNWAHNS SVEAS SNLAEPNGNERKCEPDSPVTKKMRLEPGTAAREVTNGAAHQSFGKQKSGP 1SPGHSCDLTVEGLATVTRIAPGSAKV1 D VYRTGAKCVPGAEGDSEGKPGAAPFHQVGLRVRPGRGDRTRSMSCS DKMARWNVLGCGALLMHL LEEPI YLSAVVIGKCPYSEQAMEORAL IGRCQNVSA PKFGVQELKI LQS D L LEQ SRSAVQAKR ADS PGRLVPCGAAL SWAVPEQPLDTVANGPQGTTKTI GQSLQARSQ1 SKVELFRSF QKL SRIA

RDKWPHSLRVQKLDTQOEYKEAASSYQE CAVGSRNPPDYHQFK (SEQ ID NO:72)

(underline: nucleic acid editing domain)

[0092] In some embodiments, the enzymatic domain comprises one or more of a transcriptional activator. For example, in some embodiments, the general architecture of exemplary dCas9 fusion proteins with a transcriptional activator domain comprises the structure:

\[
[NH_2]-[\text{NLS}]-[\text{Cas9}]-[(\text{transcriptional activator})_n]-[\text{COOH}],
\]

\[
[NH_2]-[\text{NLS}]-[\text{transcriptional activator}]-[\text{Cas9}]-[\text{COOH}],
\]

\[
[NH_2]-[\text{Cas9}]-[(\text{transcriptional activator})_n]-[\text{COOH}],
\]

\[
[NH_2]-[(\text{transcriptional activator})_n]-[\text{Cas9}]-[\text{COOH}];
\]

wherein NLS is a nuclear localization signal, dCas9 comprises an intein as provided herein, NH₂ is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein. In some embodiments, the fusion proteins comprises one or more repeats of the transcriptional activator, for example wherein n = 1-10 (e.g., n is 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10). In some embodiments, n = 1-20. In some embodiments, a linker is inserted between the dCas9 and the transcriptional activator domain. In some embodiments, a linker is inserted between the NLS and the transcriptional activator domain and/or dCas9 domain. In some embodiments, the NLS is located C-terminal of the transcriptional activator and/or the dCas9 domain. In some embodiments, the NLS is located between the transcriptional activator domain and the dCas9 domain. Additional features, such as sequence tags, may also be present. In some embodiments, the transcriptional activator is selected from the group consisting of VP64, (SEQ ID NO:73), VP16 (SEQ ID NO:74), and p65 (SEQ ID NO:75).

VP64

GSGRADALDDFDLMLGSDALDDFDLMLGSDALDDFDLMLGSDALDDFDLMLLin (SEQ ID NO:73)

VP16

APPTDVSLGDELHLDGEDVAMAHADALDDFDLMLGDGDSPGPGTMHDFSDAPYG GALMDMDFEF EQMFT DALKIDEYGGFEQGIR (SEQ ID NO:74)
In some embodiments, the enzymatic domain comprises one or more of a transcriptional repressor. For example, in some embodiments, the general architecture of exemplary dCas9 fusion proteins with a transcriptional repressor domain comprises the structure:

\[ \text{[NH}_2\text{]-[NLS]-[Cas9]-[transcriptional repressor]}_n\text{-[COOH]}, \]
\[ \text{[NH}_2\text{]-[NLS]-[transcriptional repressor]}_n\text{-[Cas9]-[COOH]}, \]
\[ \text{[NH}_2\text{]-[transcriptional repressor]}_n\text{-[Cas9]-[COOH]}, \]
\[ \text{[transcriptional repressor]}^\text{[transcriptional repressor]}_n\text{-[Cas9]-[COOH]}, \]

wherein NLS is a nuclear localization signal, dCas9 comprises an intein as provided herein, NH\(_2\) is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein. In some embodiments, the fusion proteins comprises one or more repeats of the transcriptional repressor, for example wherein n = 1-10 (e.g., n is 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10). In some embodiments, n = 1-20. In some embodiments, a linker is inserted between the dCas9 and the transcriptional repressor domain. In some embodiments, a linker is inserted between the NLS and the transcriptional repressor and/or dCas9 domain. In some embodiments, the NLS is located C-terminal of the transcriptional repressor and/or the dCas9 domain. In some embodiments, the NLS is located between the transcriptional repressor domain and the dCas9 domain. Additional features, such as sequence tags, may also be present. In some embodiments, the transcriptional repressor is selected from the group consisting of the KRAB (Krippel associated box) domain of Kox1, SID (mSin3 interaction domain), the CS (Chromo Shadow) domain of HP1a, or the WRPW domain of Hesl. These and other repressor domains are known in the art, and in some embodiments correspond to those described in Urrutia, KRAB-containing zinc-finger repressor proteins. *Genome Biol.* 2003;4(10):231; Gilbert et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell.* 2013; 154, 442-451; Konermann et al., Optical control of mammalian endogenous transcription and epigenetic states. *Nature.* 2013; 500, 472-476; and published U.S. patent application U.S.S.N. 14/105,017, published as U.S. 2014/0186958 Al, the entire contents of which are incorporated herein by reference. In some embodiments, the transcription repressor domain comprises one or more repeats (e.g., 2, 3, 4, 5, 6, 7, 8, 9, or 10...
repeats) of a KRAB domain. In some embodiments, the KRAB domain comprises an amino acid sequence selected from the group consisting of SEQ ID NOs:76-79. In some embodiments, the transcriptional repressor domains comprises one or more repeats of a SID protein. In some embodiments, the SID protein comprises an amino acid sequence set forth as SEQ ID NO:80. In some embodiments, the repressor domain comprises 2, 3, 4, 5, 6, 7, 8, 9, or 10 repeats of a SID protein (e.g., SEQ ID NO:80). In some embodiments, the repressor domain comprises four repeats of SID (e.g., SID4x; SEQ ID NO:81).

**KRAB** (human; GenBank: AAD20972.1)

```
MNMFKEAVTFKDVAVAFTEELGLLGPAQRKLYRDVMVENFRNLLSVGHPPFKQDVSERNEQLWIMTTATRRQGNLDPVTLPKALLLLYDLAQ
```

**KRAB protein domain**, partial (human; GenBank: CAB52478.1):

```
EQVSFKDVCVDFTQEEWYLLDPQKLYRDVI LENYSNLVS VGYCITKPEVI FKIEQGEEPWLLEKGFPSQCHP (SEQ ID NO:77)
```

**KRAB A domain**, partial (human; GenBank: AAB03530.1):

```
EAVTFKDVAVVFTEELGLLDPAQRKLYRDVMLENFRNLGS (SEQ ID NO:78)
```

**KRAB** (mouse; C2H2 type domain containing protein; GenBank: CAM27971.1):

```
MDLVTYDDVHVNFTQDEWALLDSPQKSLYKGVMLETYKNLTAIGYIWEHT1EDHFQTSRSHGSNKKT (SEQ ID NO:79)
```

**SID repressor domain**:

```
GSGMNIQMLLEAADYLERREAEHYASMLP (SEQ ID NO:80)
```

**SID4x repressor domain**:

```
GSGMNIQMLLEAADYLERREAEHYASMLPGSGMNIQMLLEAADYLERREAEHYASMLPGSMNIQMLLEAADYLERREAEHYASMLP2RSA (SEQ ID NO:81)
```

[0094] In some embodiments, the enzymatic domain comprises an epigenetic modifier or a catalytic domain thereof. For example, in some embodiments, the general architecture of exemplary dCas9 fusion proteins with an epigenetic modifier or domain comprises the structure:

```
[NH2]-[NLS]-[Cas9]-[ epigenetic modifier] -[COOH],
```
[NH₂]-[NLS]-[epigenetic modifier]-[Cas9]-[COOH],
[NH₂]-[Cas9]-[epigenetic modifier]-[COOH], or
[NH₂]-[epigenetic modifier]-[Cas9]-[COOH];

wherein NLS is a nuclear localization signal, dCas9 comprises an intein as provided herein, NH₂ is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein. In some embodiments, a linker is inserted between the dCas9 and the epigenetic modifier domain. In some embodiments, a linker is inserted between the NLS and the epigenetic modifier and/or dCas9 domain. In some embodiments, the NLS is located C-terminal of the epigenetic modifier and/or the dCas9 domain. In some embodiments, the NLS is located between the epigenetic modifier domain and the dCas9 domain. Additional features, such as sequence tags, may also be present. Epigenetic modifiers are well known in the art, and typically catalyze DNA methylation (and demethylation) or histone modifications (e.g., histone methylation/demethylation, acetylation/deacetylation, ubiquitylation, phosphorylation, sumoylation, etc.). The presence of one more epigenetic modifications can affect the transcriptional activity of one or more genes, for example turning genes from an "on" state to an "off" state, and vice versa. Epigenetic modifiers include, but are not limited to, histone demethylase, histone methyltransferase, hydroxylase, histone deacetylase, and histone acetyltransferase. Exemplary epigenetic modifying proteins can be found in Konermann et al., Optical control of mammalian endogenous transcription and epigenetic states. Nature. 2013; 500, 472-476; Mendenhall et al., Locus-specific editing of histone modifications at endogenous enhancers. Nat. Biotechnol. 2013; 31, 1133-1136; and Maeder et al., Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. Nat. Biotechnol. 2013; 31, 1137-1142; the entire contents of each are incorporated herein by reference. In some embodiments, the epigenetic modifier domain is LSD1 (Lysine (K)-specific demethylase 1A) histone demethylase, which in some embodiments, comprises in whole or in part, an amino acid sequence set forth as SEQ ID NO:82 or SEQ ID NO:83. In some embodiments, the epigenetic modifier domain is TET1 hydroxylase catalytic domain, which in some embodiments, comprises an amino acid sequence set forth as SEQ ID NO:84. In some embodiments, the epigenetic modifier is a histone deacetylase (HDAC) effector domain. In some embodiments, the HDAC effector domain comprises in whole in in part, an amino acid sequence corresponding to any of the HDAC effector proteins provided in Supplementary Table 2 of Konermann et al., Optical control of mammalian endogenous transcription and epigenetic states. Nature. 2013; 500, 472-476; SEQ ID NOs:85-96. In some embodiments, the epigenetic modifier is a histone
methyltransferase (HMT) effector domain. In some embodiments, the HMT effector domain comprises in whole or in part, an amino acid sequence corresponding to any of the HDAC effector proteins provided in Supplementary Table 3 of Konermann et al., Optical control of mammalian endogenous transcription and epigenetic states. Nature. 2013; 500, 472-476; SEQ ID NOs:97-106.

LSD1, isoform a (human):
MLSGKKAAAAAASAAAAATGTEAPGTAAGSNGSEVAAPAGLSGPAEVPGAPAVGERTPKKEPRA
SPPGPLAEPGSAQPQAGPTVPGPATMPETGIAETPEGRTSSRKRRAKVEYREMDESANLSEDEYY
SEEERNAKAEKKEKLLPPPPQPAPPEEENEESEPEPSGVGVEAAFSRPLQHDMTSQEAAFCFPDI
ISGPQQTQKVFLFIRNRLQLMLNPKIQLTFEATLQQLEAPYNSNDTVALVHRVH
SHTERYRLNIFGIYKRKLPPTKKTGKVIIGGSGVSLAAARQLQSFVGDVTLEARDRVGRVATF
RKNYNADLGAMVTGLGNMFAVSKQVNMELAKIKQCPFLYEANGQAVPKEDEMVEQEFNFR
LLELYHSLHDFNVLNKNPVSGALEVVIQLKEHASHKVEDEQIEHWWKIKVTQETELKLENNKNVNLK
EKHLQHQQJEASEVPPRDITAEFLVSKHDRLTALCAYDELAETQGKLEKIALQDELEANPSDVY
LSSRDRQILDWHFANLEFANPLTSLSKHWDQDDDFETGSGLTMRNGYSCVPALEAGLDIKILN
AVRQVRYTASGCEIVAIVNTRTSQTFIXYKCDVCTLCPVGLKQPPAVQFPPLPEWKTSAVRQMGF
GNLKVLCFDRFDVFDPVSNLQGVTSGTASRGEFLFWNLYYKAPILLAVGAEAAGIMENI
SDDVIVGRCAILKIGFSSAVQPKETVSSRWRADFPMARGVSYVSAAGSSGNYDLMQPITGPSIPGAPQP
IPRLFFAGEHTIRNPATVHGALLSGLREALGRIADQFLGAMYTLPRQATPGVPAQPSM (SEQ ID NO:X)

LSD1, isoform b (human):
MLSGKKAAAAAASAAAAATGTEAPGTAAGSNGSEVAAPAGLSGPAEVPGAPAVGERTPKKEPRA
SPPGPLAEPGSAQPQAGPTVPGPATMPETGIAETPEGRTSSRKRRAKVEYREMDESANLSEDEYY
SEEERNAKAEKKEKLLPPPPQPAPPEEENEESEPEPSGVGVEAAFSRPLQHDMTSQEAAFCFPDI
ISGPQQTQKVFLFIRNRLQLMLNPKIQLTFEATLQQLEAPYNSNDTVALVHRVH
SHTERYRLNIFGIYKRKLPPTKKTGKVIIGGSGVSLAAARQLQSFVGDVTLEARDRVGRVATF
RKNYNADLGAMVTGLGNMFAVSKQVNMELAKIKQCPFLYEANGQAVPKEDEMVEQEFNFR
LLELYHSLHDFNVLNKNPVSGALEVVIQLKEHASHKVEDEQIEHWWKIKVTQETELKLENNKNVNLK
EKHLQHQQJEASEVPPRDITAEFLVSKHDRLTALCAYDELAETQGKLEKIALQDELEANPSDVY
LSSRDRQILDWHFANLEFANPLTSLSKHWDQDDDFETGSGLTMRNGYSCVPALEAGLDIKILN
AVRQVRYTASGCEIVAIVNTRTSQTFIXYKCDVCTLCPVGLKQPPAVQFPPLPEWKTSAVRQMGF
GNLKVLCFDRFDVFDPVSNLQGVTSGTASRGEFLFWNLYYKAPILLAVGAEAAGIMENI
SDDVIVGRCAILKIGFSSAVQPKETVSSRWRADFPMARGVSYVSAAGSSGNYDLMQPITGPSIPGAPQP
IPRLFFAGEHTIRNPATVHGALLSGLREALGRIADQFLGAMYTLPRQATPGVPAQPSM (SEQ ID NO:X)

TET1 catalytic domain:
SIVAQLSRPDALAAALNTNVLALACLGRPGPAVKGLPHAPALIKRTNRIPEFSTSRHVADHAQV
VRVLGFFQCHSHPAQADDAMQFGMSGGSNPCPSCLDVRQKVKGPFYTHLGAQPSVAAREIMEN
RYQGKGNAIRIEIVYVTGEKKSCHHGCPlVKLWRSDEEKVLCLVRQGTHCRIPTMVLMILWMD
GIPLMFADLRYLETLENKYSHNHDRTDRTNLNRRTCTCIDPETGCASTSFSGCWSMYNCGKFG
RPSRSPRFRIDPSLPHELKNEDLPLATLAPIKXYAPVAYNQVYENERVACRILSKEGRPFS
GVTCADLFCAHPRHTHHMRNNSGTVCTLREETDRSGLVIPQDELHVLPLLYKSLDSTDEFGSKGEMA
KIKSGAEVLPARRRKRRTCTQFEPVKSKKAAAMTEVLAKHAVKKEPIPRKRRNNTTNTNNSKP
SSLPTLGSNTETVQPEKVSETFKILKSDDNKTYSLMPASPAPVKEASGFSWSKPKSTASATAPLKN
DATAASCNFSDRPSTHPCTMPSRGLSANAGADSNIGGQDIQVGYLAPLSAPNLIPEIAHEPSSDPL
SDDLPSPAEEKLPHIDEYWSDEHIFLDANIIGVAPAHGSLVIECARRELHATTVEHPNHRNTHPRLSLVFQYKHKLKNPFQHGFELNK
FEAKEAKNKKMKASEQKDQAANEGPEQSSEVNELNQIPSHKALTLLTHDNVTTVSPYALTHVAGFYNHW
v (SEQ ID NO:X)

HP AC effector domains:

HDAC8 (*X. laevis*):

ASSPKKKRKVEASMSRVKVASMEEMAADFHTDAYLQQLHKVSEEGDNDDPETLEYGLGYDCPITEG
IYDYAAAAGGATLTAEESEQKIGKTR AVNWPGGWGHAHKKDEAGSFCLYNDAVLGILKREKDRVLYV
DMQLHGHGFVEDASFTSKMVTSHLKFSPGFFPGTGDSDIGLGKGYRYSINVPDQDGIDQKYQI
CEGVKLKFTSFNEAVVLQGADTI AGDPMCSFNMTPEGIKCGLKYVLQWPQTLILGGGYPHLPNT
ARCWYTILALTIVGRTLSEIPDHEFFTEGPDYVLEITPSRNDNTKVQEVILQSIKGNLKRVVEF
(SEQ ID NO:85)

RPD3 (*S. cerevisiae*):

ASSPKKKRKVEASRVAYFYDADVNGYAYGAGHMPKHRVRMAHSLIMNYGLYKMEI YRAKPATKQE
MCQFHTDEIDFSRLVTPDNLMKERESVFNVDGDCVFDGLYEYCE ISGGGSMGAARLNRGKCDV
AVNYAGGLHAHKKASEAGSFCLYNCLVI GI IEELHYRFRVLIDIVHGGDVEEAFYTTDRVMTCSF
H KYGREFPFTGTGELRDIGVGAGKNYAVERNVRDIDATYRVSFEPVKKIMEWYQPASAVVLQCGGDSL
GDLRCFNSMAMCHVNYKSFGIPFMVVGGGGYTMVRNVAATWCFETGLLNVLDKLPYE F
(SEQ ID NO:86)

MesoLo4 (*M. loti*):

ASSPKKKRKVEASPMQI VHPHDYGAFATNHRFMSKYPULLMEALRAGLPALDNLTTEPAPASWL
KLAHAAAYVDQVISCEVEKIEREGFVPPVGLNLAQALTGTILAAALRALRIG ACNTAAGSHHAR
RAGAGGFCFTNDAVAHLVVLDGAIQNLIRCVLDVGDDGTADILSDEPGVTFSMHGERNYPVK
ASDLDDIALPDTGDAAYLRRATILPELSARARMDI VFYNAVGVDVAEDRLGLSLNGGLRDEMV
IGHFRALEGIPVGVGGYSTDVPALASRHAILFEVASTYAEF (SEQ ID NO:87)

HDAC11 (human):

ASSPKKKRKVEASMLHTTQLYQHPETRWPIVYSPRYNITFMGELEHLHPFDAGKWKGVINFLKEKLL
SDMSLVEAREASEEIDLVYVHRMLKESWFPAVATITEIPVFILNPFLVRKLRPLRTQTGTGIM
AGKLAVERGSGNGFHCSSGCAVDEADITAFLFEVGERISATI IDLDAHGNGHERD
FMDKRVYIMDVHNYHYPGRFAKQRAIRKVELEWTDEDELDKVERNKKSLQHELPDVVYNAG
TDLEGDRLGLSI SAGVIRKDRLFVDRGMVRPVLPMLYSGYKQTARI IADS ILNLFGLGLIGP
ESPSVSAQNSDTPPLPPAVPEF (SEQ ID NO:88)

HDT1 (*A. thaliana*):

ASSPKKKRKVEASMEFVGIEVKGPKVTPTEGILIHVSQAASGLGECNKKGEFVPLHVKVGNQNVL
GTLSTENIPQLOFCDLVFDEKEFELSHTGKGSYFVYKTFPNIEPQGYSEEEDDEEEdFPAGNAAKAVA
KPKAKAEVKPAVDEEESDSMDGEDDSGDEEEDEEETPKPAPPSSKRNANETTFKAPVSAKAKV
AVTPQKTDKEKKGGKAANQSEF (SEQ ID NO:89)

SIRT3 (human):

ASSPKKKRKVEASMSVGASTIPSGIFDPFRSPGSGSINQLQQYDLPYPEAIFELFFFHNPKPFLLLAK
ELYPNYKNPTNYLFLRLLHDKGLLRLYTNQIDGERSVGIPASKLVEAHTGATATCTVCQRFPFG
EDIRADVMDRVPJRCPCTGVKRPVVFGEFELPQRLLLHYVDFPMADLDDILGLTGSLVEFPFASLTEA
VRSSVFLIRLNRDLVGPALWHPRSRDVAQLGDVVGVHGVSLVELLWTEERMDLQRETGKLGDGPKEF
(SEQ ID NO:90)
HST2 (*S. cerevisiae*):
ASSPKKKRKVEASTEMSVRKIAAHMKSNPNAKVIFMVGAGISTSCGIPDFRSPGTGLYHNLARLKLPHYPEAVFDFDQFSDFLPFYTikalEYPGNFRPSKFHYLLKLFQDKVLKVYTQNIIDTLERQAGVKDDLIEAEHSGFAHCHICCGKVKPQFVFKSKLAEHP1KDFVKCVDGELVKPAIVFGEGLpDPSFSETWLNDESWLREKITTSGKHQPQPLVIVVGTSSLAVYPFASLPEEIIPRKRVLNLETVGDFANKRPDDTLIVHQYSDEFAEQLVEELGWQDDEFELKITAQQGMEF (SEQ ID NO:91)

CobB (*E. coli* (*K12*)):
ASSPKKKRKVEASMEKPRVLVTGAGISAEVTGIRTFRAADGLWEEHRVEDVATPEGFDRPDELVQAFYNNARRRQEQIPQNPAANHLALAKQLDALGDFLVLVTQNIDLNHERAGNTNVHMGELLKVCRCSQSGQVLDWTGDVTPEDCKCCQFPAPLRPHVVFHELPMGLMEIYMALSMDIFIAIGTSHVYPAAFGVHCKLHGAHTVLENLEPSQVNGEFAEKYYGPASQVPEFVEKLLKGLKAGSEIF (SEQ ID NO:92)

HST2 (*C. albicans*):
ASSPKKKRKVEASMPSLDDILKPVAEAVNKGGKVTFFNGAGISTGAGIPDFRSPDTGLYANLAKLNLPFAEAVFDFIDFKEFKYPYTLAEELYPNFLAPFTKFHPFIIKLQDQGLKVRVYTQNIDTLERGVEDKIVEAHGSFASNHCVDHEMTETLKTYMDDKIPSCQHCEVYKPVDFIVFEGREPVKFDFDLWEDDCEDVEVAVAGSTLVIDFAPAFLPGENVKKCLRVNLNKEVGTFFHEPRKSIJA1HDCDIAVERLCTTLDGLDKLNEVYKEKIKYSKATEIKMHEIEDKLLKEAHLKEDKHTVKKLIKNDANDKEEQLIDKAKAEF (SEQ ID NO:93)

SIRT5 (human):
ASSPKKKRKVEASSSMADFRKKFKAHKVIISGAGVSAESGVTFRFARGGYYWRRQWQADLATPLAFAHNPSRVWEFYHYRREVMGKSPEPNAGHRIAECETRLGKQGRRVVTQNIEDLRHKAGTKNLEIHGLFKTRCTSGVAENYKPSICPALSAGKAGEPPTQDASIPVEKLPREEAGCGGGLRPVVFENGENDPAIIEEVEDRAHCDLCLCGTSSYVPPAMFPAQVAPAEFTETTPATNRFHFFQPCGCTTLPEALACHENETVSEF (SEQ ID NO:94)

Sir2A (*P. falciparum*):
ASSPKKKRKVEASMGNLMISFLKDTQSIITELEAKIHKCKHVVALTGSSTSAESNIPSFRGSSNSISWSYDPR1YGTIGWFWKYPEKIMEVIREIDSDYEIEIEINGHVALSTLESGLYKSVTNQVNGGLHEASGNTKVISSHGNVEAFVCCCTNKIVKLQMQLKTSHPHMQLFPECPCGIFPKNIIFEGVSSDLKEEIEIAKCDLLLVIGTSTSYSTANLCHFACKKKIKVIENISKTYTNKMSDYHVCAKFSLTVAEIGKSSKEKNKIMEF (SEQ ID NO:95)

SIRT6 (human):
ASSPKKKRKVEASMVNYAAGNPSYADCGKCLPHELIDPEELELRKVEWELARLWQSSSVVFHTGAGISTAGIPDFRPGPHWTMEERLAKPFTTFTESARTQHTMALVQLRERVGLRFLVSNQVNGGLHRSFPRKDLAEHLHNMVEECAKCTQYVRDTVGTMGLKTAGRLCRAICRLDRTLDWEDSLPDRDLALADEASNRANDLSTILGTSSLQHSGNPLATPRRHGRLVTVNQLPTKHHRADLRHIHGVDHMVTMLKHLGIEIPAWDGRPVLERALPPEF (SEQ ID NO:96)

HMT effector domains:

NUE (*C. trachomatis*):
ASSPKKKRKVEASMTTNSTQLYLSLHGGIDSAIPYPVRRVEQLQFSLPFLQFNAAVQKRIQRLCYREEKLAVLSSLAKWLQHGLKQLRMAPKNNPVAICWINSYVYGVFAREISIPAMSYIGETGTIBQRALWLDENDYCFRYPYRPSFRTIDSGMQGNTRFINHSNPNLEIAIGAFENGIPHIIRAIKDILPGEEELCYHGYLPYWKHHKREEFVPQEEEF (SEQ ID NO:97)
vSET (P. bursaria chlorella virus):
ASSPKKKRKVEASMFQHFKQDLEREELLRRHRSTPRHLDPSLANLYLVKAKQRRAARLEWQ
ELNAKRSHLRIVVYVDEVDDPGFPPAFVYNEYRVEGEITLNLQNAVVCQCDLWAPTPGCAGSLH
KFAYNDDQGVRALQLPLYECPNCRCGCDPCVRKVQKVGTVLEDILKIRKNSF
VMEVGEITSEEAARRQIYDSQGATYLFVLDLEDYTVDAAAYGNISHFVNVHCDPNLQVYVNFIDNLDERLPRAAFFATRITARGEELTFDYMNQVDPVDMESTRMDSNFGLAQLGPLPSKKVRIEECKGTE
SCRKYLFEF (SEQ ID NO:98)

SUV39H1 (human):
ASSPKKKRKVEASNLKCVRILKQFHKDLERELLRRHHRSKTPRHLPSLANLYLVKAKQRRAARLEWQ
ELNAKRSHLRIVVYVDEVDDPGFPPAFVYNEYRVEGEITLNLQNAVVCQCDLWAPTPGCAGSLH
KFAYNDDQGVRALQLPLYECPNCRCGCDPCVRKVQKVGTVLEDILKIRKNSF
VMEVGEITSEEAARRQIYDSQGATYLFVLDLEDYTVDAAAYGNISHFVNVHCDPNLQVYVNFIDNLDERLPRAAFFATRITARGEELTFDYMNQVDPVDMESTRMDSNFGLAQLGPLPSKKVRIEECKGTE
SCRKYLFEF (SEQ ID NO:98)

DIM5 (N. crassa):
ASSPKKKRKVEASMEKAFRPHFFHNHGDPANPEKKKNCWCHCQIRSFATHAQFLPI
SIVNREDDAFLNPNFRFIDSHISIKGNVPMVADQSFVRGCSCASDEECMYSTCQLDEMAPDSDEEAPTYTRKRFAYYSQGAK
KGLLRDRLQSQEPIYECHQGCACSKCDPCNVRVegerTQPLQFRTKDRGVKCPVKNRQFVQDRY
LGEIITSEEAARRRRAESTIARRDVYFALDKFSDFPSLPLNLQQMPELDEGYEMSGPTRFPINHCDP
MAIFARVGDHADKHIDDLALFAIKDIPGCTELTFDYMNQVDPVDMESTRMDSNFGLAQLGPLPSKKVRIEECKGTE
SCRKYLFEF (SEQ ID NO:98)

KYP (A. thaliana):
ASSPKKKRKVEASISGGEFKIGPAPNTRVDDVSPSTSGFYTIKSLIIEPNVIIPSSTGNCNCRGSC
TSDKKCAAKLNGNFPYVDLNGLRIEESRDFVFECPGHCCGPCKVNRISQKRLFNLLEVFRSASKG
WAVMRYEYIPAAGPSVCEYIGVVRADVTISDNEYIIFIEDQCTMTQGLGQRQRLDQVADWPMNGV
QSSEDENAPFCICADGSGNARFHUCPENLQVFVCVLSHDQFILRVALVLFADNIPMQLETLYDY
GIALDVSHEF (SEQ ID NO:100)

SUVR4 (A. thaliana):
ASSPKKKRKVEASIGLYFLKIGPAPNTRVDDVSPSTSGFYTIKSLIEPNVIIPSSTGCNCLSGADDPCNCTCARETSGEYAYTKEGILLKEKFL
DTCLKMKKEPDSFKVKYCDPLERTDHDGTKYAGCDGHLILRIKFIKCEWRCNKCDMOCNVRVQGQRC
QLQVYFTQEGKGWLRNLQDPKFTICEYHGEILINTELYDRNVRRSSERHYPVTVLADWSGKDNL
KDEEAALCDATIGCNVARFINSRCEDANMIDIEIEPTDFHYHAAFFTLRDVKAMEELTDWYIDFN
DKSHPVKAFRCGCGSERCDRKIKGSQGKSERRKIVSAKKQGSKEVSKKRF (SEQ ID NO:102)

Set4 (C. elegans):
ASSPKKKRKVEASMQLHEQIQANISVTFNDLPRSDHMTPLTELCDYFDATLTLVDVQNFTTHKMSKK
RRYYLYQDEYRTARVTMVKEQDWDRTNAINYGLLTLRSVSHEFSKLPNNKLFERFRDHIVRFILNMFLDS
GYTIQECEKRYSEQGHQKALVSTGVSRYGDKIERLSGVVCCILSSEDEDSILAQEGSDFSVYMYSTRKRC
STILWLGPAAYNHDCPTECVFVHSGTHAISIRRVLDMDVPGDEITCFYGEFFGPNIDCECTCECKMN
GAFSYLRGBENAEPIIESKTKYELRSREF (SEQ ID NO:103)

Set1 (C. elegans):
ASSPKKKRKVEASMKVAKLTSRMKDRCAAASPPSDIENSEPSLASHSSSSGMRTPSNRSCR
KGVSXKDIVSNHKTIFEFQVRNRNRTKSWQSEDAAKHELRTDLTVGLGNERLLEVYKVDGKGIRTVNK
FEKGDFVYEYGRVMMESEAKVIEQYNSDIEEGYIMGYFENKHCCIDATEKESPWKGRLINHSVLR
PNLKTKVVEIDGSLHILLVARRQIAQGEELLYDYDGRSAETIAKNPWLVNTF (SEQ ID NO:104)
In some embodiments, ligand-dependent Cas9-intein variants are provided herein that exhibit decreased off-target activity. For example, in some embodiment, Cas9-intein variants are provided herein that comprise a Cas9 nuclease domain, or a nuclease-deficient Cas9 domain and a heterologous nucleic acid-editing domain, such as, for example, a heterologous nuclease domain, a recombinase domain, or a deaminase domain. In some such embodiments, the ligand-dependent Cas9-inteins provided herein exhibit decreased, minimal, or no off-target activity in the presence of a ligand at a concentration effective to effect excision of the intein from the Cas9-intein variant, or at a concentration effective to induce a desired modification (e.g., cleavage, nicking, recombination, or deamination) of a target site. In some embodiments, the ligand-dependent Cas9-intein variants provided herein exhibit an off-target activity in their active state (e.g., in the presence of or after being contacted with a suitable ligand) that is decreased as compared to the off-target activity of wild-type Cas9. For example, in some embodiments, the off-target activity of a Cas9-intein variant is decreased to less than 80%, less than 75%, less than 50%, less than 45%, less than 35%, less than 30%, less than 25%, less than 24%, less than 23%, less than 22%, less than 21%, less than 20, less than 19%, less than 18%, less than 17%, less than 16%, less than 15%, less than 14%, less than 13%, less than 12%, less than 11%, less than 10%, less than 9%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, or less than 1% of wild-type Cas9 under the same conditions.

Pharmaceutical compositions

In some embodiments, any of the ligand-dependent site-specific enzymes described herein are provided as part of a pharmaceutical composition. For example, some embodiments provide pharmaceutical compositions comprising a Cas9 protein comprising an intein, or fusion proteins comprising a dCas9 protein with an intein fused to a nuclease,
recombinase, deaminase, or a transcriptional activator as provided herein, or a nucleic acid encoding such a protein, and a pharmaceutically acceptable excipient. Pharmaceutical compositions may further comprise one or more gRNA(s).

[0097] In some embodiments, compositions provided herein are administered to a subject, for example, to a human subject, in order to effect a targeted genomic modification within the subject. In some embodiments, cells are obtained from the subject and are contacted with an inventive ligand-dependent site-specific enzyme ex vivo. In some embodiments, cells removed from a subject and contacted ex vivo with an inventive ligand-dependent site-specific enzyme are re-introduced into the subject, optionally after the desired genomic modification has been effected and/or detected in the cells. Although the descriptions of pharmaceutical compositions provided herein are principally directed to pharmaceutical compositions which are suitable for administration to humans, it will be understood by the skilled artisan that such compositions are generally suitable for administration to animals of all sorts. Modification of pharmaceutical compositions suitable for administration to humans in order to render the compositions suitable for administration to various animals is well understood, and the ordinarily skilled veterinary pharmacologist can design and/or perform such modification with merely ordinary, if any, experimentation. Subjects to which administration of the pharmaceutical compositions is contemplated include, but are not limited to, humans and/or other primates; mammals, domesticated animals, pets, and commercially relevant mammals such as cattle, pigs, horses, sheep, cats, dogs, mice, and/or rats; and/or birds, including commercially relevant birds such as chickens, ducks, geese, and/or turkeys.

[0098] Formulations of the pharmaceutical compositions described herein may be prepared by any method known or hereafter developed in the art of pharmacology. In general, such preparatory methods include the step of bringing the active ingredient into association with an excipient, and then, if necessary and/or desirable, shaping and/or packaging the product into a desired single- or multi-dose unit.

[0099] Pharmaceutical formulations may additionally comprise a pharmaceutically acceptable excipient, which, as used herein, includes any and all solvents, dispersion media, diluents, or other liquid vehicles, dispersion or suspension aids, surface active agents, isotonic agents, thickening or emulsifying agents, preservatives, solid binders, lubricants and the like, as suited to the particular dosage form desired. Remington's *The Science and Practice of Pharmacy*, 21st Edition, A. R. Gennaro (Lippincott, Williams & Wilkins, Baltimore, MD, 2006; incorporated in its entirety herein by reference) discloses various excipients used in
formulating pharmaceutical compositions and known techniques for the preparation thereof. See also PCT application PCT/US20 10/055 131, incorporated in its entirety herein by reference, for additional suitable methods, reagents, excipients and solvents for producing pharmaceutical compositions comprising a nuclease. Except insofar as any conventional excipient medium is incompatible with a substance or its derivatives, such as by producing any undesirable biological effect or otherwise interacting in a deleterious manner with any other component(s) of the pharmaceutical composition, its use is contemplated to be within the scope of this disclosure.

In some embodiments, compositions in accordance with the present invention may be used for treatment of any of a variety of diseases, disorders, and/or conditions, including, but not limited to, autoimmune disorders (e.g. diabetes, lupus, multiple sclerosis, psoriasis, rheumatoid arthritis); inflammatory disorders (e.g. arthritis, pelvic inflammatory disease); infectious diseases (e.g. viral infections (e.g., HIV, HCV, RSV), bacterial infections, fungal infections, sepsis); neurological disorders (e.g. Alzheimer's disease, Huntington's disease; autism; Duchenne muscular dystrophy); cardiovascular disorders (e.g. atherosclerosis, hypercholesterolemia, thrombosis, clotting disorders, angiogenic disorders such as macular degeneration); proliferative disorders (e.g. cancer, benign neoplasms); respiratory disorders (e.g. chronic obstructive pulmonary disease); gastrointestinal disorders (e.g. inflammatory bowel disease, ulcers); musculoskeletal disorders (e.g. fibromyalgia, arthritis); endocrine, metabolic, and nutritional disorders (e.g. diabetes, osteoporosis); genitourinary disorders (e.g. renal disease); psychological disorders (e.g. depression, schizophrenia); skin disorders (e.g. wounds, eczema); and blood and lymphatic disorders (e.g. anemia, hemophilia); etc.

Methods

In another aspect of this disclosure, methods for site-specific nucleic acid (e.g., DNA) modification are provided. In some embodiments, the methods comprise contacting a DNA with any of the ligand-dependent Cas9 proteins (complexed with a gRNA) described herein, either before or after contacting the protein with a ligand that induces self-excision of the ligand-dependent intein thereby activating the nuclease. For example, in some embodiments, the method comprises (a) contacting a RNA-guided nuclease (e.g., a Cas9 protein including Cas9 nickase) comprising a ligand-dependent intein with a ligand, wherein binding of the ligand to the intein induces self-excision of the intein; and (b); contacting a DNA with the RNA-guided nuclease, wherein the RNA-guided nuclease is
associated with a gRNA; whereby self-excision of the intein from the RNA-guided nuclease in step (a) allows the RNA-guided nuclease to cleave the DNA, thereby producing cleaved DNA. In some embodiments, for examples those involving the use of an intein containing Cas9 nickase, the method produces a single strand break in the DNA. In some embodiments, the method produces a double strand break in the DNA. In some embodiments, the RNA-guided nuclease is able to bind a gRNA and optionally bind a target nucleic acid prior to being contacted with a ligand that induces self-excision of the intein, but the RNA-guided nuclease is unable to cleave the target nucleic acid until self-excision of the intein occurs. In some embodiments, the RNA-guided nuclease is unable to bind a gRNA and therefore is unable to bind a target nucleic acid until the RNA-guided nuclease is contacted with a ligand that induces self-excision of the intein. In some embodiments, the RNA-guided nuclease is any nuclease comprising Cas9 (or a variant or a fragment thereof) which comprises a ligand-dependent intein as provided herein.

[00102] In some embodiments, the method involves the use of fusion proteins comprising a nuclease-inactivated Cas9 (e.g., dCas9) fused to a nuclease domain (e.g., FokI), wherein the fusion protein comprises a ligand-dependent intein (e.g., in the dCas9 domain as provided herein), and the fusion protein lacks one or more activities (as described herein) prior to excision of the intein. In some embodiments, the fusion protein is any fusion protein described herein. In some embodiments, the method comprises contacting a target nucleic acid (e.g., DNA) with two such fusion proteins, each comprising a distinct gRNA that targets the nucleic acid, and the method comprises contacting the target nucleic acid with two such fusion proteins. The method increases the specificity of cleavage, and therefore decreases off target effects, as two fusions are required to bind the target site to elicit any nuclease activity as the nuclease domains fused to the dCas9 domain typically must dimerize at the target site to induce cleavage. In some embodiments, the method comprises contacting the fusion proteins with a ligand that induces self-excision of the intein, either before or after the gRNAs bind the fusion proteins, and/or before or after the fusion proteins bind the target nucleic acid. Once the fusion proteins are activated following excision of the intein, the nuclease domains (e.g., the FokI domains) dimerize and cleave and the target nucleic acid. Compositions and methods of using dCas9-FokI fusions are known to those of skill in the art (see, e.g., U.S. Patent Application No.: 14/320,498, titled "CAS9 VARIANTS AND USES THEREOF" which was filed on June 30, 2014; the entire contents of which are incorporated herein by reference). Those of skill in the art are routinely able to design appropriate gRNAs that target two of the fusion proteins to a target nucleic acid, and understand that in some
aspects the gRNAs are designed to hybridize to regions of the target nucleic acid that are no more than 10, no more than 15, no more than 20, no more than 25, no more than 30, no more than 40, no more than 50, no more than 60, no more than 70, no more than 80, no more than 90, or no more than 100 base pairs apart.

[00103] In some embodiments, any of the methods provided herein can be performed on DNA in a cell. For example, in some embodiments the DNA contacted by any of the inventive ligand-dependent site-specific Cas9 enzymes provided herein is in a eukaryotic cell. In some embodiments, the eukaryotic cell is in an individual. In some embodiments, the individual is a human. In some embodiments, any of the methods provided herein are performed in vitro. In some embodiments, any of the methods provided herein are performed in vivo.

[00104] In some embodiments of this disclosure, methods for site-specific nucleic acid (e.g., DNA) recombination are provided. In some embodiments, the methods are useful for inducing recombination of or between two or more regions of two or more nucleic acids (e.g., DNA). In some embodiments, the methods are useful for inducing recombination of or between two or more regions in a single nucleic acid molecule (e.g., DNA). Because the recombination fusion proteins used in the methods are ligand-dependent, the timing of recombination can be controlled to minimize off-target effects. In some embodiments, the recombination of one or more target nucleic acid molecules requires the formation of a tetrameric complex at the target site. Typically, the tetramer comprises four (4) inventive RNA-guided recombinase fusion proteins (e.g., a complex of any four inventive recombinase fusion proteins provided herein). In some embodiments, each recombinase fusion protein of the tetramer targets a particular DNA sequence via a distinct gRNA bound to each recombinase fusion protein. In some embodiments, the fusion proteins are first contacted with a ligand that induces self-excision of the intein, thereby allowing the fusion proteins to (i) bind a gRNA, (ii) bind a target nucleic acid(s), and (iii) form a complex to induce recombination between the target nucleic acid(s). In some embodiments, the fusion proteins are able to bind a gRNA prior to excision of the intein and optionally are able to bind the target nucleic acid(s) but are unable to induce recombination until the intein is excised (e.g., through the addition of a ligand that binds the ligand-dependent intein). Any of the ligand-dependent recombinase fusion proteins provided herein are useful for methods for site-specific recombination.

[00105] In some embodiments, the method for site-specific recombination between two DNA molecules comprises (a) contacting a first DNA with a first ligand-dependent
RNA-guided recombinase fusion protein, wherein the nuclease-inactivated Cas9 domain
binds a first gRNA that hybridizes to a region of the first DNA; (b) contacting the first DNA
with a second ligand-dependent RNA-guided recombinase fusion protein, wherein the
nuclease-inactivated Cas9 domain of the second fusion protein binds a second gRNA that
hybridizes to a second region of the first DNA; (c) contacting a second DNA with a third
ligand-dependent RNA-guided recombinase fusion protein, wherein the nuclease-inactivated
Cas9 domain of the third fusion protein binds a third gRNA that hybridizes to a region of the
second DNA; and (d) contacting the second DNA with a fourth ligand-dependent RNA-
guided recombinase fusion protein, wherein the nuclease-inactivated Cas9 domain of the
fourth fusion protein binds a fourth gRNA that hybridizes to a second region of the second
DNA. In some embodiments, the fusion proteins are first contacted with a ligand that
induces self-excision of the intein prior to forming a complex with a gRNA and/or prior to
hybridizing with a target DNA. In some embodiments, the method comprises contacting the
fusion proteins with the ligand after the fusion proteins form a complex and/or hybridizes to a
target DNA. Typically, the binding of the fusion proteins in steps (a) - (d) results in the
tetramerization of the recombinase catalytic domains of the fusion proteins, such that the
DNAs are recombined (i.e., following excision of the intein). In some embodiments, the
gRNAs of steps (a) and (b) hybridize to opposing strands of the first DNA, and the gRNAs of
steps (c) and (d) hybridize to opposing strands of the second DNA. In some embodiments,
the target sites of the gRNAs of steps (a) - (d) are spaced to allow for tetramerization of the
recombinase catalytic domains. For example, in some embodiments, the target sites of the
gRNAs of steps (a) - (d) are no more than 10, no more 15, no more than 20, no more than 25,
no more than 30, no more than 40, no more than 50, no more than 60, no more than 70, no
more than 80, no more than 90, or no more than 100 base pairs apart. In some embodiments,
the two regions of the two DNA molecules being recombined share homology, such that the
regions being recombined are at least 80%, at least 90%, at least 95%, at least 98%, or are
100% homologous.

[00106] In some embodiments, methods for site-specific recombination between two
regions of a single DNA molecule are provided. In some embodiments, the methods
comprise (a) contacting a DNA with a first dCas9-recombinase fusion protein, wherein the
dCas9 domain binds a first gRNA that hybridizes to a region of the DNA; (b) contacting the
DNA with a second dCas9-recombinase fusion protein, wherein the dCas9 domain of the
second fusion protein binds a second gRNA that hybridizes to a second region of the DNA;
(c) contacting the DNA with a third dCas9-recombinase fusion protein, wherein the dCas9
domain of the third fusion protein binds a third gRNA that hybridizes to a third region of the DNA; and (d) contacting the DNA with a fourth dCas9-recombinase fusion protein, wherein the dCas9 domain of the fourth fusion protein binds a fourth gRNA that hybridizes to a fourth region of the DNA. In some embodiments, the fusion proteins are first contacted with a ligand that induces self-excision of the intein prior to forming a complex with a gRNA and/or prior to hybridizing with a target DNA. In some embodiments, the method comprises contacting the fusion proteins with the ligand after the fusion proteins form a complex and/or hybridizes to a target DNA. Typically, the binding of the fusion proteins in steps (a) - (d) results in the tetramerization of the recombinase catalytic domains of the fusion proteins, such that the DNA is recombined (e.g. following the excision of the intein). In some embodiments, two of the gRNAs of steps (a) - (d) hybridize to the same strand of the DNA, and the other two gRNAs of steps (a)- (d) hybridize to the opposing strand of the DNA. In some embodiments, the gRNAs of steps (a) and (b) hybridize to regions of the DNA that are no more 10, no more than 15, no more than 20, no more than 25, no more than 30, no more than 40, no more than 50, no more than 60, no more than 70, no more than 80, no more than 90, or no more than 100 base pairs apart, and the gRNAs of steps (c) and (d) hybridize to regions of the DNA that are no more than 10, no more 15, no more than 20, no more than 25, no more than 30, no more than 40, no more than 50, no more than 60, no more than 70, no more than 80, no more than 90, or no more than 100 base pairs apart. In some embodiments, the two regions of the DNA molecule being recombined share homology, such that the regions being recombined are at least 80%, at least 90%, at least 95%, at least 98%, or are 100% homologous.

[00107] In some embodiments, any of the inventive methods for site-specific recombination are amenable for inducing recombination, such that the recombination results in excision (e.g., a segment of DNA is excised from a target DNA molecule), insertion (e.g., a segment of DNA is inserted into a target DNA molecule), inversion (e.g., a segment of DNA is inverted in a target DNA molecule), or translocation (e.g., the exchange of DNA segments between one or more target DNA molecule(s)). In some embodiments, the particular recombination event (e.g., excision, insertion, inversion, translocation, etc.) depends, inter alia, on the orientation (e.g., with respect to the target DNA molecule(s)) of the bound RNA-guided recombinase fusion protein(s). In some embodiments, the orientation, or direction, in which a RNA-guided recombinase fusion protein binds a target nucleic acid can be controlled, e.g., by the particular sequence of the gRNA bound to the RNA-guided recombinase fusion protein(s). Methods for controlling or directing a particular
recombination event are known in the art, and include, for example, those described by Turan and Bode, "Site-specific recombinases: from tag-and-target- to tag-and-exchange-based genomic modifications." FASEB J. 2011; Dec;25(12):4088-107, the entire contents of which are hereby incorporated by reference.

[00108] In some embodiments, any of the methods for site-specific recombination can be performed in vivo or in vitro. In some embodiments, any of the methods for site-specific recombination are performed in a cell (e.g., recombining genomic DNA in a cell). The cell can be prokaryotic or eukaryotic. The cell, such as a eukaryotic cell, can be in an individual, such as a subject, as described herein (e.g., a human subject). The methods described herein are useful for the genetic modification of cells in vitro and in vivo, for example, in the context of the generation of transgenic cells, cell lines, or animals, or in the alteration of genomic sequence, e.g., the correction of a genetic defect, in a cell in or obtained from a subject. In some embodiments, a cell obtained from a subject and modified according to the methods provided herein, is re-introduced into a subject (e.g., the same subject), e.g., to treat a disease, or for the production of genetically modified organisms in agricultural, medical, or biological research.

[00109] In applications in which it is desirable to recombine two or more nucleic acids so as to insert a nucleic acid sequence into a target nucleic acid, a nucleic acid comprising a donor sequence to be inserted is also provided, e.g., to a cell. By a "donor sequence" it is meant a nucleic acid sequence to be inserted at the target site induced by one or more RNA-guided recombinase fusion protein(s). In some embodiments, e.g., in the context of genomic modifications, the donor sequence will share homology to a genomic sequence at the target site, e.g., 1%, 2%, 3%, 4%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, or 100% homology with the nucleotide sequences flanking the target site, e.g., within about 100 bases or less of the target site, e.g. within about 90 bases, within about 80 bases, within about 70 bases, within about 60 bases, within about 50 bases, within about 40 bases, within about 30 bases, within about 15 bases, within about 10 bases, within about 5 bases, or immediately flanking the target site. In some embodiments, the donor sequence does not share any homology with the target nucleic acid, e.g., does not share homology to a genomic sequence at the target site. Donor sequences can be of any length, e.g., 10 nucleotides or more, 50 nucleotides or more, 100 nucleotides or more, 250 nucleotides or more, 500 nucleotides or more, 1000 nucleotides or more, 5000 nucleotides or more, 10000 nucleotides or more, 100000 nucleotides or more, etc.
[00110] Typically, the donor sequence is not identical to the target sequence that it replaces or is inserted into. In some embodiments, the donor sequence contains at least one or more single base changes, insertions, deletions, inversions, or rearrangements with respect to the target sequence (e.g., target genomic sequence). In some embodiments, donor sequences also comprise a vector backbone containing sequences that are not homologous to the DNA region of interest and that are not intended for insertion into the DNA region of interest.

[00111] The donor sequence may comprise certain sequence differences as compared to the target (e.g., genomic) sequence, for example, restriction sites, nucleotide polymorphisms, selectable markers (e.g., drug resistance genes, fluorescent proteins, enzymes etc.), which can be used to assess successful insertion of the donor sequence at the target site or in some cases may be used for other purposes (e.g., to signify expression at the targeted genomic locus). In some embodiments, if located in a coding region, such nucleotide sequence differences will not change the amino acid sequence, or will make silent amino acid changes (e.g., changes which do not affect the structure or function of the protein). In some embodiments, these sequences differences may include flanking recombination sequences such as FLPs, loxP sequences, or the like, that can be activated at a later time for removal of e.g., a marker sequence. The donor sequence may be provided to the cell as single-stranded DNA, single-stranded RNA, double-stranded DNA, or double-stranded RNA. It may be introduced into a cell in linear or circular form. If introduced in linear form, the ends of the donor sequence may be protected (e.g., from exonucleolytic degradation) by methods known to those of skill in the art. For example, one or more dideoxynucleotide residues are added to the 3’ terminus of a linear molecule and/or self-complementary oligonucleotides are ligated to one or both ends. See, e.g., Chang et al., Proc. Natl. Acad Sci USA. 1987; 84:4959-4963; Nehls et al, Science. 1996; 272:886-889. In some embodiments, a donor sequence is introduced into a cell as part of a vector molecule having additional sequences such as, for example, replication origins, promoters, and genes encoding antibiotic resistance. In some embodiments, donor sequences can be introduced as naked nucleic acid, as nucleic acid complexed with an agent such as a liposome or polymer (e.g., poloxamer), or can be delivered by viruses (e.g., adenovirus, AAV, etc.).

[00112] In some embodiments, any of the methods provided herein can be performed on DNA in a cell. For example, in some embodiments the DNA contacted by any RNA/gRNA-comprising complex provided herein is in a eukaryotic cell. In some embodiments, the eukaryotic cell is in an individual. In some embodiments, the individual is
In some embodiments, any of the methods provided herein are performed in vitro. In some embodiments, any of the methods provided herein are performed in vivo.

[00113] In some embodiments of this disclosure, methods for site-specific nucleic acid editing are provided. In some embodiments, the fusion protein is used to introduce a point mutation into a nucleic acid by deaminating a target nucleobase, e.g., a cytidine (C) residue. In some embodiments, the method comprises contacting a DNA molecule with a ligand-dependent fusion protein comprising a nuclease inactivated RNA-guided nuclease (e.g., dCas9), which comprises a ligand dependent intein, fused to a deaminase, and (b) a gRNA targeting the fusion protein of step (a) to a target nucleotide sequence of the DNA strand; wherein the DNA molecule is contacted with the fusion protein and the gRNA in an amount effective and under conditions suitable for the deamination of a nucleotide base. Any of the fusion proteins comprising a gene editing domain as provided herein are amenable for use in the methods. In some embodiments, the method first comprises contacting the fusion protein with a ligand that induces self-excision of the intein prior to forming a complex with the gRNA. In some embodiments, the method comprises contacting the fusion protein with a ligand that induces self-excision of the intein after the fusion protein has formed a complex with the gRNA.

[00114] In some embodiments, the deamination of the target nucleobase results in the correction of a genetic defect, e.g., in the correction of a point mutation that leads to a loss of function in a gene product. In some embodiments, the genetic defect is associated with a disease or disorder, e.g., a lysosomal storage disorder or a metabolic disease, such as, for example, type I diabetes. In some embodiments, the methods provided herein are used to introduce a deactivating point mutation into a gene or allele that encodes a gene product that is associated with a disease or disorder. For example, in some embodiments, methods are provided herein that employ a DNA editing fusion protein to introduce a deactivating point mutation into an oncogene (e.g., in the treatment of a proliferative disease). A deactivating mutation may, in some embodiments, generate a premature stop codon in a coding sequence, which results in the expression of a truncated gene product, e.g., a truncated protein lacking the function of the full-length protein.

[00115] In some embodiments, the purpose of the methods provided herein is to restore the function of a dysfunctional gene via genome editing. Compositions and methods of using gene editing enzymes fused e.g., to Cas9 are known, and include those described in U.S. Patent Application No.: 14/325,815 titled "FUSIONS OF CAS9 DOMAINS AND NUCLEIC ACID-EDITING DOMAINS," and filed on July 8, 2014; the entire contents of
which are incorporated herein by reference. The fusion proteins provided herein (comprising ligand-dependent inteins) can be validated for gene editing-based human therapeutics in vitro, e.g., by correcting a disease-associated mutation in human cell culture. It will be understood by the skilled artisan that the fusion proteins provided herein, e.g., the fusion proteins comprising a dCas9 domain (e.g., comprising a ligand-dependent intein) and a nucleic acid deaminase domain can be used to correct any single point T -> C or A -> G mutation. In the first case, deamination of the mutant C back to U corrects the mutation, and in the latter case, deamination of the C that is base-paired with the mutant G, followed by a round of replication, corrects the mutation.

[00116] An exemplary disease-relevant mutation that can be corrected by the provided fusion proteins in vitro or in vivo is the H1047R (A3140G) polymorphism in the PI3KCA protein. The phosphoinositide-3-kinase, catalytic alpha subunit (PI3KCA) protein acts to phosphorylate the 3-OH group of the inositol ring of phosphatidylinositol. The PI3KCA gene has been found to be mutated in many different carcinomas, and thus it is considered to be a potent oncogene (Lee et al., PIK3CA gene is frequently mutated in breast carcinomas and hepatocellular carcinomas. Oncogene. 2005; 24(8): 1477-80). In fact, the A3140G mutation is present in several NCT60 cancer cell lines, such as, for example, the HCT1 16, SKOV3, and T47D cell lines, which are readily available from the American Type Culture Collection (ATCC)(Ikediobi et al., Mutation analysis of 24 known cancer genes in the NCT60 cell line set. Mol Cancer Ther. 2006; 5(11):2606-12.

[00117] In some embodiments, a cell carrying a mutation to be corrected, e.g., a cell carrying a point mutation, e.g., an A3140G point mutation in exon 20 of the PI3KCA gene, resulting in a H1047R substitution in the PI3KCA protein is contacted with an expression construct encoding a ligand-dependent Cas9 deaminase fusion protein and an appropriately designed gRNA targeting the fusion protein to the respective mutation site in the encoding PI3KCA gene. Control experiments can be performed where the gRNAs are designed to target the fusion enzymes to non-C residues that are within the PI3KCA gene. Genomic DNA of the treated cells can be extracted, and the relevant sequence of the PI3KCA genes PCR amplified and sequenced to assess the activities of the fusion proteins in human cell culture.

[00118] It will be understood that the example of correcting point mutations in PI3KCA is provided for illustration purposes and is not meant to limit the instant disclosure. The skilled artisan will understand that the instantly disclosed ligand-dependent DNA-editing fusion proteins can be used to correct other point mutations and mutations associated with other cancers and with diseases other than cancer including other proliferative diseases.
[00119] The successful correction of point mutations in disease-associated genes and alleles opens up new strategies for gene correction with applications in therapeutics and basic research. Site-specific single-base modification systems like the disclosed fusions of ligand-dependent Cas9 and deaminase enzymes or domains also have applications in "reverse" gene therapy, where certain gene functions are purposely suppressed or abolished. In these cases, site-specifically mutating Trp (TGG), Gin (CAA and CAG), or Arg (CGA) residues to premature stop codons (TAA, TAG, TGA) can be used to abolish protein function in vitro, ex vivo, or in vivo.

[00120] The instant disclosure provides methods for the treatment of a subject diagnosed with a disease associated with or caused by a point mutation that can be corrected by a ligand-dependent Cas9 DNA editing fusion protein provided herein. For example, in some embodiments, a method is provided that comprises administering to a subject having such a disease, e.g., a cancer associated with a PI3KCA point mutation as described above, an effective amount of a ligand-dependent Cas9 deaminase fusion protein that corrects the point mutation or introduces a deactivating mutation into the disease-associated gene, e.g., following subsequent administration of the small molecule (e.g., ligand) that activates the fusion protein. In some embodiments, the disease is a proliferative disease. In some embodiments, the disease is a genetic disease. In some embodiments, the disease is a neoplastic disease. In some embodiments, the disease is a metabolic disease. In some embodiments, the disease is a lysosomal storage disease. Other diseases that can be treated by correcting a point mutation or introducing a deactivating mutation into a disease-associated gene will be known to those of skill in the art, and the disclosure is not limited in this respect.

[00121] The instant disclosure provides methods for the treatment of additional diseases or disorders, e.g., diseases or disorders that are associated or caused by a point mutation that can be corrected by deaminase-mediated gene editing. Some such diseases are described herein, and additional suitable diseases that can be treated with the strategies and fusion proteins provided herein will be apparent to those of skill in the art based on the instant disclosure. Exemplary suitable diseases and disorders include, without limitation, cystic fibrosis (see, e.g., Schwank et al., Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. Cell stem cell. 2013; 13: 653-658; and Wu et al., Correction of a genetic disease in mouse via use of CRISPR-Cas9. Cell stem cell. 2013; 13: 659-662, neither of which uses a deaminase fusion protein to correct the genetic defect); phenylketonuria - e.g., phenylalanine to serine mutation at position 835 in
phenylalanine hydroxylase gene (T>C mutation) - see, e.g., McDonald et al., *Genomics*. 1997; 39:402-405; Bernard-Soulier syndrome (BSS) - e.g., phenylalanine to serine mutation at position 55 in the platelet membrane glycoprotein IX (T>C mutation) - see, e.g., Noris et al., *British Journal of Haematology*. 1997; 97: 312-320; epidermolytic hyperkeratosis (EHK) - e.g., leucine to proline mutation at position 160 in keratin 1 (T>C mutation) - see, e.g., Chipev et al., *Cell*. 1992; 70: 821-828; chronic obstructive pulmonary disease (COPD) - e.g., leucine to proline mutation at position 55 in α1-antitrypsin (T>C mutation) - see, e.g., Poller et al., *Genomics*. 1993; 17: 740-743; Charcot-Marie-Tooth disease type 4J - e.g., leucine to proline mutation at position 197 in FIG4 (T>C mutation) - see, e.g., Kundu et al., *3 Biotech*. 2013; 3: 225-234; neuroblastoma (NB) - e.g., isoleucine to threonine mutation at position 509 in von Willebrand disease (vWD) - e.g., cysteine to arginine mutation at position 509 in von Willebrand factor (T>C mutation) - see, e.g., Lavergne et al., *Br. J. Haematol.*. 1992; 82: 66-72; myotonia congenital - e.g., cysteine to arginine mutation at position 277 in the muscle chloride channel gene CLCN1 (T>C mutation) - see, e.g., Weinberger et al., *The J. of Physiology*. 2012; 590: 3449-3464; hereditary renal amyloidosis - e.g., stop codon to arginine mutation at position 78 in apolipoprotein AII (T>C mutation) - see, e.g., Yazaki et al., *Kidney Int.*. 2003; 64: 11-16; dilated cardiomyopathy (DCM) - e.g., tryptophan to Arginine mutation at position 148 in the FOXD4 gene (T>C mutation), see, e.g., Minoretti et al., *Int. J. of Mol. Med.*. 2007; 19: 369-372; hereditary lymphedema - e.g., histidine to arginine mutation at position 1035 in tyrosine kinase (A>G mutation), see, e.g., Irthum et al., *Am. J. Hum. Genet.*. 2000; 67: 295-301; familial Alzheimer's disease - e.g., isoleucine to valine mutation at position 143 in presenilin1 (A>G mutation), see, e.g., Gallo et al., *J. Alzheimer’ s disease*. 2011; 25: 425-431; Prion disease - e.g., methionine to valine mutation at position 129 in prion protein (A>G mutation) - see, e.g., Lewis et al., *J. of General Virology*. 2006; 87: 2443-2449; chronic infantile neurologic cutaneous articular syndrome (CINCA) - e.g., Tyrosine to Cysteine mutation at position 570 in cryopyrin (A>G mutation) - see, e.g., Fujisawa et al. *Blood*. 2007; 109: 2903-2911; and desmin-related myopathy (DRM) - e.g., arginine to glycine mutation at position 120 in αB crystallin (A>G mutation) - see, e.g., Kumar et al., *J. Biol. Chem.*. 1999; 274: 24137-24141. The entire contents of each of the foregoing references and database entries are incorporated herein by reference.

According to another aspect, methods for transcriptional activation of a gene are provided. In some embodiments, the methods comprise contacting a DNA molecule
comprising a gene with (a) a ligand-dependent dCas9 fusion protein comprising a transcriptional activator (e.g., any of those provided herein) and (b) a gRNA targeting the fusion protein of (a) to a target nucleotide sequence of the DNA strand; wherein the DNA molecule is contacted with the fusion protein and the gRNA in an amount effective and under conditions suitable for the transcriptional activation of the gene. In some embodiments, the method further comprises contacting the fusion protein with a ligand that induces self-excision of the intein. In some embodiments, the fusion protein is contacted with the ligand prior to forming a complex with a gRNA. In some embodiments, the fusion protein is contacted with the ligand after forming a complex with a gRNA. In some embodiments, the gRNA targets the promoter region of a gene. Methods for inducing gene activation using fusion proteins comprising a transcriptional activator are known in the art, and include those described by Perex-Pinera et al., "RNA-guided gene activation by CRISPR-Cas9-based transcription factors." Nature Methods. 2013; 10, 973-976; the entire contents of which are incorporated herein by reference. [00123] According to another aspect, methods for transcriptional repression of a gene are provided. In some embodiments, the methods comprise contacting a DNA molecule comprising a gene with (a) a ligand-dependent dCas9 fusion protein comprising a transcriptional repressor (e.g., any of those provided herein) and (b) a gRNA targeting the fusion protein of (a) to a target nucleotide sequence of the DNA strand; wherein the DNA molecule is contacted with the fusion protein and the gRNA in an amount effective and under conditions suitable for the transcriptional repression of the gene. In some embodiments, the method further comprises contacting the fusion protein with a ligand that induces self-excision of the intein. In some embodiments, the fusion protein is contacted with the ligand prior to forming a complex with a gRNA. In some embodiments, the fusion protein is contacted with the ligand after forming a complex with a gRNA. In some embodiments, the gRNA targets the promoter region of a gene. Methods for inducing gene repression using fusion proteins comprising a transcriptional repressor are known in the art, and include those described by Gilbert et al., CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell. 2013; 154, 442-451; the entire contents of which are incorporated herein by reference. [00124] According to another aspect, methods for epigenetic modification of DNA are provided. In some embodiments, the methods comprise contacting a DNA molecule comprising with (a) a ligand-dependent dCas9 fusion protein comprising an epigenetic modifier (e.g., any of those provided herein) and (b) a gRNA targeting the fusion protein of
(a) to a target nucleotide sequence of the DNA strand; wherein the DNA molecule is contacted with the fusion protein and the gRNA in an amount effective and under conditions suitable for the epigenetic modification of the DNA. In some embodiments, the DNA comprises one or more genes. In some embodiments, the method further comprises contacting the fusion protein with a ligand that induces self-excision of the intein. In some embodiments, the fusion protein is contacted with the ligand prior to forming a complex with a gRNA. In some embodiments, the fusion protein is contacted with the ligand after forming a complex with a gRNA. In some embodiments, the gRNA targets the promoter region of a gene. In some embodiments, the epigenetic modification that results is methylation of DNA. In some embodiments, the epigenetic modification that results is demethylation of DNA. In some embodiments, the epigenetic modification that results is methylation of histone protein. In some embodiments, the epigenetic modification that results is demethylation of histone protein. In some embodiments, the epigenetic modification that results is acetylation of histone protein. In some embodiments, the epigenetic modification that results is deacetylation of histone protein. Methods for inducing epigenetic modifications using fusion proteins comprising an epigenetic modifier are known in the art, and include those described by Konermann et al., Optical control of mammalian endogenous transcription and epigenetic states. Nature. 2013; 500, 472-476; Mendenhall et al., Locus-specific editing of histone modifications at endogenous enhancers. Nat. Biotechnol. 2013; 31, 1133-1136; and Maeder et al., Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. Nat. Biotechnol. 2013; 31, 1137-1142; the entire contents of which are incorporated herein by reference.

In some embodiments, any of the methods provided herein can be performed on DNA in a cell. For example, in some embodiments the DNA contacted by any RNA/gRNA-comprising complex provided herein is in a eukaryotic cell. In some embodiments, the eukaryotic cell is in an individual. In some embodiments, the individual is a human. In some embodiments, any of the methods provided herein are performed in vitro. In some embodiments, any of the methods provided herein are performed in vivo.

In some embodiments of the methods provided herein, the ligand-dependent Cas9 protein, e.g., the Cas9-intein or the Cas9-intein fusion protein, is contacted with the ligand at a concentration effective to excise the intein from the Cas9-intein variant, or at a concentration effective to induce a desired modification (e.g., cleavage, nicking, recombination, or deamination) of a target site. In some embodiments, a ligand-dependent
Cas9 protein provided herein is contacted with a suitable ligand at a concentration resulting in decreased off-target activity of the Cas9 protein as compared to the off-target activity of wild-type Cas9. For example, in some embodiments, a method provided herein comprises contacting a population of ligand-dependent Cas9 proteins in vitro or in vivo in the presence of a target nucleic acid to be modified with a suitable ligand at a concentration resulting in the desired modification of the target nucleic acid, and in either no off-target activity (i.e., no modification of any non-target nucleic acids) or in an off-target activity of less than 80%, less than 75%, less than 50%, less than 45%, less than 40%, less than 35%, less than 30%, less than 25%, less than 24%, less than 23%, less than 22%, less than 21%, less than 20, less than 19%, less than 18%, less than 17%, less than 16%, less than 15%, less than 14%, less than 13%, less than 12%, less than 11%, less than 10%, less than 9%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, or less than 1% of the off-target activity observed or expected under the same conditions when using wild-type Cas9.

Polynucleotides, Vectors, Cells, Kits

[00127] In another aspect of this disclosure, polynucleotides encoding one or more of the inventive proteins and/or gRNAs are provided. For example, polynucleotides encoding any of the proteins described herein are provided, e.g., for recombinant expression and purification of isolated nucleases, recombinases, gene editing enzymes, and other nucleic acid modifying enzymes, e.g., comprising Cas9 variants (e.g., dCas9) comprising ligand-dependent inteins. In some embodiments, an isolated polynucleotide comprises one or more sequences encoding a ligand dependent RNA-guided nuclease (e.g., Cas9). In some embodiments, an isolated polynucleotide comprises one or more sequences encoding a Cas9 fusion protein, for example, any of the Cas9 fusion proteins described herein (e.g., those comprising a nuclease-inactivated Cas9 fused to a nuclease, recombinase, deaminase domain, or transcriptional activator). In some embodiments, an isolated polynucleotides comprises one or more sequences encoding a gRNA, alone or in combination with a sequence encoding any of the proteins described herein.

[00128] In some embodiments, vectors encoding any of the proteins described herein are provided, e.g., for recombinant expression and purification of Cas9 proteins, and/or fusions comprising Cas9 proteins (e.g., variants). In some embodiments, the vector comprises or is engineered to include an isolated polynucleotide, e.g., those described herein. In some embodiments, the vector comprises one or more sequences encoding a Cas9 protein
(as described herein), a gRNA, or combinations thereof, as described herein. Typically, the vector comprises a sequence encoding an inventive protein operably linked to a promoter, such that the fusion protein is expressed in a host cell.

[00129] In some embodiments, cells are provided, e.g., for recombinant expression and purification of any of the Cas9 proteins provided herein. The cells include any cell suitable for recombinant protein expression, for example, cells comprising a genetic construct expressing or capable of expressing an inventive protein (e.g., cells that have been transformed with one or more vectors described herein, or cells having genomic modifications, for example, those that express a protein provided herein from an allele that has been incorporated into the cell's genome). Methods for transforming cells, genetically modifying cells, and expressing genes and proteins in such cells are well known in the art, and include those provided by, for example, Green and Sambrook, Molecular Cloning: A Laboratory Manual (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)) and Friedman and Rossi, Gene Transfer: Delivery and Expression of DNA and RNA, A Laboratory Manual (1st ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2006)).

[00130] Some aspects of this disclosure provide kits comprising a ligand-dependent Cas9 variant (e.g., a ligand dependent Cas9 nuclease (or nickase), and/or a ligand-dependent dCas9 variant fused to a nuclease, recombinase, deaminase, or a transcriptional activator as provided herein. In some embodiments, the kit comprises a polynucleotide encoding an inventive Cas9 variant, nuclease, recombinase, and/or deaminase e.g., as provided herein. In some embodiments, the kit comprises a vector for recombinant protein expression, wherein the vector comprises a polynucleotide encoding any of the proteins provided herein. In some embodiments, the kit comprises a cell (e.g., any cell suitable for expressing Cas9 proteins or fusions comprising Cas9 proteins, such as bacterial, yeast, or mammalian cells) that comprises a genetic construct for expressing any of the proteins provided herein. In some embodiments, any of the kits provided herein further comprise one or more gRNAs and/or vectors for expressing one or more gRNAs. In some embodiments, the kit comprises an excipient and instructions for contacting the Cas9 proteins or dCas9 fusions with the excipient to generate a composition suitable for contacting a nucleic acid with the inventive protein. In some embodiments, the composition is suitable for delivering an inventive protein to a cell, or for delivering a nucleic acid encoding the protein to a cell. In some embodiments, the composition is suitable for delivering an inventive protein to a subject. In some embodiments, the excipient is a pharmaceutically acceptable excipient.
EXEMPLARY

**Example 1: Small molecule-controlled Cas9**

Cas9 variants that can be activated in the presence of a small molecule were engineered, allowing spatiotemporal control over DNA cleavage. These engineered Cas9 variants contain a small-molecule-regulated intein (Buskirk *et al.*, *Proc. Natl. Acad. Sci. USA*. 2004; **101**, 10505-10510), which has been optimized for mammalian cells (Peck *et al.*, *Chem. Biol*. 2011; **18**(5), 619-630), that renders the protein inactive as a nuclease. Upon addition of the cell-permeable molecule, 4-hydroxytamoxifen (4-HT), the intein excises itself from the protein and ligates the flanking extein sequences, restoring Cas9 activity. Because these Cas9 variants can be active over a smaller time window than wild-type Cas9, the likelihood of having off-target cleavage is reduced.

The 37R3-2 intein was inserted at 15 different positions into human codon-optimized *Streptococcus pyogenes* Cas9 (e.g., SEQ ID NO:2). The intein was inserted in place of a single cysteine, alanine, serine, or threonine residue. Upon excision, the intein leaves a cysteine residue. Thus, the primary structure generated following protein splicing is either identical to the unmodified version of Cas9 when the intein is inserted in place of cysteine, or it is one amino acid different when the intein is inserted in place of alanine, serine, or threonine.

Plasmid constructs were generated in which the intein replaced amino acid residues: Cys80, Ala127, Thr146, Ser219, Thr333, Thr519, Cys574, Thr622, Ser701, Ala728, Thr995, Ser1006, Ser1154, Ser1159, or Ser1274 (e.g., in the amino acid sequence set forth as SEQ ID NO:2). These plasmids express the Cas9 variant with a nuclear localization signal (NLS) and 3xFLAG tag from the CMV promoter.

HEK293-GFP stable cells were transfected with the Cas9 expression plasmid, a gRNA (targeting Emerald GFP; Guilinger *et al.*, *Nature Biotechnology* (2014)), and iRFP670 (transfection control), using Lipofectamine 2000. Twelve hours after transfection, media, either containing 4-HT (1 µM) or without 4-HT, was added.

Five days after transfection, cells were trypsinized and analyzed on a flow cytometer. Cells lacking GFP indicated genome modification. Cas9 variants that induced minimal genome modification in the absence of 4-HT but induce significant genome modification in the presence of 4-HT were deemed small-molecule-regulated variants in this Example. Of fifteen targeted insertions, five demonstrated minimal genome modification in the absence of 4-HT. These variants are highlighted in bold in the Table I below.
Additionally, a time course was performed in which incubation with 4-HT was limited to 2, 4, 8, 12 or 24 hours, after which point the media was replaced. Presumably, the shorter time an “active” cas9 is present, the less off-target cleavage. As depicted in Table II below, treating with 4-HT for 2 hours is sufficient for on-target cleavage and longer treatment periods do not show significant increased cleavage in this assay.

To assess the ability of the ligand-dependent Cas9 proteins to affect genomic modifications in the presence of absence of ligand, HEK293-GFP stable cells (GenTarget) were transfected with Cas9 expression plasmids and sgRNAs targeting EMX, VEGF, or CLTA genomic sites using Lipofectamine 2000 as previously described (Guilinger et al., Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification Nature Biotechnology. 2014; 32(6):577-82). 4-HT (1 μM) was added during transfection for + 4-HT samples. 12 hours after transfection the media was replaced. 60 hours after transfection, cells were trypsinized and genomic DNA was isolated using the DNAAdvance kit (Agencourt). 40-80 ng of genomic DNA was used as a template to PCR amplify the targeted genomic loci with flanking Survey primer pairs as previously described (Guilinger et al., Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification Nature Biotechnology. 2014; 32(6):577-82). PCR products were purified with a QIAquick PCR Purification Kit (Qiagen) and quantified with a Quant-iT PicoGreen dsDNA Kit (Life Technologies). 200 ng of purified PCR DNA was then combined with 2 μL of NEBuffer 2 (NEB) in a total volume of 19 μL, and denatured then re-annealed with thermocycling at 95°C for 5 min, 95-85°C at 2°C/s, 85-20°C at 0.2°C/s. The re-annealed DNA was incubated with 1 μL of T7 Endonuclease I (10U/μL, NEB) at 37°C for 15 min. 10 μL of 50% glycerol was added to the T7 Endonuclease reaction and 15 μL was analyzed on a 5% TBE 18-well Criterion PAGE gel (Bio-Rad) electrophoresed for 30 min at 200V and stained with EtBr for 15 min.

As shown in Figure 2, the addition of 4-HT to ligand-dependent Cas9:Intein variants (Cas9:Intein with 37R3-2 replacing S219 (SEQ ID NO:30) and Cas9:Intein with 37R3-2 replacing C574 (SEQ ID NO:33)) resulted in genomic modification of the target sites, comparable to modification by wild-type Cas9. In the absence of 4-HT, the Cas9:Intein variants displayed minimal or no modification of the EMX and VEGF genomic target sites, while some background cleavage was observed for the CLTA genomic target site. Gene modification levels can be estimated by comparing the intensities of the cleaved (two smaller fragments) and uncleaved bands. These results demonstrate that Cas9 cleavage of genomic
target sites can be controlled by the addition of ligand (here, 4-HT) which activates the proteins.

Table 1

<table>
<thead>
<tr>
<th>Cas9 Variant</th>
<th>Cells without GFP (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- 4-HT</td>
</tr>
<tr>
<td>None</td>
<td>4.65</td>
</tr>
<tr>
<td>wild-type cas9</td>
<td>48.49</td>
</tr>
<tr>
<td>intein(Cys80)-Cas9 (SEQ ID NO:27)</td>
<td>7.08</td>
</tr>
<tr>
<td>intein(Ala127)-Cas9 (SEQ ID NO:28)</td>
<td><strong>7.97</strong></td>
</tr>
<tr>
<td>intein(Thr146)-Cas9 (SEQ ID NO:29)</td>
<td><strong>8.77</strong></td>
</tr>
<tr>
<td>intein(Ser219)-Cas9 (SEQ ID NO:30)</td>
<td><strong>6.53</strong></td>
</tr>
<tr>
<td>intein(Thr333)-Cas9 (SEQ ID NO:31)</td>
<td>4.96</td>
</tr>
<tr>
<td>intein(Thr519)-Cas9 (SEQ ID NO:32)</td>
<td><strong>9.49</strong></td>
</tr>
<tr>
<td>intein(Cys574)-Cas9 (SEQ ID NO:33)</td>
<td><strong>5.74</strong></td>
</tr>
<tr>
<td>intein(Thr622)-Cas9 (SEQ ID NO:34)</td>
<td>5.67</td>
</tr>
<tr>
<td>intein(Ser701)-Cas9 (SEQ ID NO:35)</td>
<td>6.54</td>
</tr>
<tr>
<td>intein(Ala728)-Cas9 (SEQ ID NO:36)</td>
<td>20.82</td>
</tr>
<tr>
<td>intein(Thr995)-Cas9 (SEQ ID NO:37)</td>
<td>14.95</td>
</tr>
<tr>
<td>intein(Ser1006)-Cas9 (SEQ ID NO:38)</td>
<td>6.80</td>
</tr>
<tr>
<td>intein(Ser1154)-Cas9 (SEQ ID NO:39)</td>
<td>21.14</td>
</tr>
<tr>
<td>intein(Ser1159)-Cas9 (SEQ ID NO:40)</td>
<td>5.65</td>
</tr>
<tr>
<td>intein(Ser1274)-Cas9 (SEQ ID NO:41)</td>
<td>3.08</td>
</tr>
</tbody>
</table>
Table 2

<table>
<thead>
<tr>
<th>Cas9 Variant</th>
<th>Cells without GFP (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- 4-HT</td>
</tr>
<tr>
<td></td>
<td>2 hrs</td>
</tr>
<tr>
<td>None</td>
<td>2.98</td>
</tr>
<tr>
<td>wild-type cas9</td>
<td>34.65</td>
</tr>
<tr>
<td>intein(Ala127)-Cas9 (SEQ ID NO:28)</td>
<td>5.03</td>
</tr>
<tr>
<td>intein(Thr146)-Cas9 (SEQ ID NO:29)</td>
<td>4.28</td>
</tr>
<tr>
<td>intein(Ser219)-Cas9 (SEQ ID NO:30)</td>
<td>3.92</td>
</tr>
<tr>
<td>intein(Thr519)-Cas9 (SEQ ID NO:32)</td>
<td>4.29</td>
</tr>
<tr>
<td>intein(Cys574)-Cas9 (SEQ ID NO:33)</td>
<td>2.91</td>
</tr>
</tbody>
</table>

Example 2: Small Molecule-controlled Cas9 Protein with Improved Genome-Editing Specificity

[00139] Cas9 nucleases that are activated by the presence of a cell-permeable small molecule were developed by inserting an evolved 4-hydroxytamoxifen (4-HT)-responsive intein at specific positions in Cas9. In human cells, conditionally active Cas9s modify target genomic sites with up to 25-fold higher specificity than wild-type Cas9.

[00140] The RNA-guided endonuclease Cas9 from the type II CRISPR-Cas system enables simple and efficient genome editing in a wide variety of organisms. Virtually any target DNA locus can be cleaved by programming Cas9 with a single guide RNA (sgRNA) that contains a stretch of 10–20 nucleotides complementary to the target sequence. Due to its simplicity and robustness, the Cas9 system has been widely adopted for biological research and therapeutic development. The DNA cleavage specificity of Cas9 is imperfect, however, raising concerns over off-target genome modification that may limit its usefulness in therapeutic or research applications. Cas9 off-target activity has been reduced through
protein\textsuperscript{9-12} and sgRNA\textsuperscript{13} engineering, and by direct delivery of Cas9:sgRNA protein:RNA complexes into cells\textsuperscript{14-16}.

A complementary, underexplored strategy to improve Cas9 specificity is to reduce its activity once it has had sufficient opportunity to modify the target DNA locus. Indeed, higher concentrations of Cas9 in cells have been observed to degrade specificity\textsuperscript{4,6} (defined as the ratio of on-target:off-target DNA cleavage activity), presumably because any Cas9 protein present after the target locus has been modified can only process off-target substrates. Unfortunately, wild-type Cas9 nucleases are not known to be regulated by other molecules and therefore are used in constitutively active form. While Cas9 can be regulated at the transcriptional level through the use of inducible promoters\textsuperscript{17,18}, transcriptional control cannot limit activity to the short temporal windows that may be necessary to maximize genome-editing specificity\textsuperscript{16,19}, in contrast with the high temporal resolution of post-translational strategies that directly control protein activity.

Engineered variants of Cas9 that can be controlled with a readily available, cell-permeable small molecule were developed. We previously evolved inteins that undergo protein splicing only in the presence of 4-hydroxytamoxifen (4-HT)\textsuperscript{20}. These inteins were developed by inserting the human estrogen receptor ligand-binding domain into the \textit{M. tuberculosis} RecA intein and evolving the resulting inactive fusion protein into a conditionally active intein that requires the presence of 4-HT\textsuperscript{20-22}. Subsequent evolution at 37°C yielded a second-generation intein, 37R3-2, with improved splicing properties in mammalian cells\textsuperscript{22}. We envisioned that inserting the 37R3-2 intein into Cas9 at a location that disrupts Cas9 activity until protein splicing has taken place could result in conditionally active Cas9 nucleases that are active only in the presence of 4-HT (Fig. 3a).

We genetically inserted the 4-HT-dependent intein at each of fifteen positions in Cas9 (Cys80, AlaL27, ThrL46, Ser219, Thr333, Thr519, Cys574, Thr622, Ser701, AlaL728, Thr995, Serl006, SerL54, Serl59, and Serl274), chosen to distribute the location of the intein across the structural domains of Cas9\textsuperscript{23} (Fig. 3b and Example 1). Because intein splicing leaves behind a single Cys residue, the intein was inserted in place of one Cas9 amino acid in each of the 15 candidate constructs. In addition to replacing natural Cys amino acids, we also favored replacing Ala, Ser, or Thr residues to minimize the likelihood that the resulting Cys point mutation resulting from protein splicing would disrupt Cas9 activity. The 15 intein-Cas9 candidates were expressed in HEK293-GFP cells together with a sgRNA that targets the genomic \textit{EGFP} locus in these cells. Twelve hours post-transfection, cells were treated with or without 1 \(\mu\text{M} \) 4-HT. Five days post-transfection, cells were analyzed on a
flow cytometer for loss of GFP expression from Cas9-mediated EGFP cleavage and subsequent non-homologous end joining.

[00144] Eight of the candidates, corresponding to intein insertion at A127, T146, S219, T333, T519, C574, S1006, and S1159, demonstrated 4-HT-dependent loss of GFP expression consistent with 4-HT-triggered Cas9 activity (Fig. 3c). Interestingly, three intein-Cas9 proteins (insertion at A728, T995, and S1154) showed high DNA modification rates both in the presence and absence of 4-HT, suggesting that large protein insertions at these positions do not significantly inhibit nuclease activity, or that the intein lost its 4-HT dependence due to context-dependent conformational perturbations. We speculate that it may be possible to engineer split Cas9 variants by dividing the protein at these locations, given their tolerance of a 413-residue insertion. The lack of nuclease activity of the remaining four Cas9-inteins (insertion at C80, T622, S701, and S1274) in the presence or absence of 4-HT could result from the inability of the intein to splice in those contexts, the inability of Cas9 to refold properly following splicing, or intolerance of replacement of native Thr or Ser residues with Cys. We pursued two intein-Cas9 variants corresponding to insertion at S219 and C574 (Fig. 3b). These two variants combined high activity in the presence of 4-HT and low activity in the absence of 4-HT.

[00145] To evaluate the genome modification specificity of conditionally active Cas9 variants, we expressed intein-Cas9(S219), intein-Cas9(C574), and wild-type Cas9 in HEK293-GFP cells together with each of three previously described sgRNAs that target the well-studied EMX, VEGF, and CLTA genomic loci. We assayed these Cas9:sgRNA combinations in human cells for their ability to modify the three on-target loci as well as 11 known off-target genomic sites (Table 3). Cells were treated with or without 1 \( \mu \)M 4-HT during transfection, and after 12 h the media was replaced with fresh media lacking 4-HT. We observed no cellular toxicity arising from 12 or 60 h of treatment with 1 \( \mu \)M 4-HT in untransfected or transfected HEK293 cells (Fig. 5). Genomic DNA was isolated 60 h post-transfection and analyzed by high-throughput DNA sequencing.

[00146] Overall on-target genome modification frequency of intein-Cas9(S219) and intein-Cas9 (C574) in the presence of 1 \( \mu \)M 4-HT was similar to that of wild-type Cas9 (Fig. 4a, Tables 4 and 5). On-target modification frequency in the presence of 4-HT was 3.4- to 7.3-fold higher for intein-Cas9(S219), and 3.6- to 9.6-fold higher for intein-Cas9(C574), than in the absence of 4-HT, whereas modification efficiency for wild-type Cas9 was 1.2- to 1.8-fold lower in the presence of 4-HT (Fig. 4a). Both intein-Cas9 variants exhibited a low level of background activity in the absence of 4-HT, consistent with previous reports. Western
blot analysis of intein-Cas9(S219) from transfected HEK293 cells confirmed the presence of spliced product at the earliest assayed time point (4 h) following 4-HT treatment; no spliced product was detected in the absence of 4-HT (Fig. 6). Together, these results indicate that intein-Cas9(S219) and intein-Cas9(C574) are slightly less active than wild-type Cas9 in the presence of 4-HT, likely due to incomplete splicing (Fig. 6), but much less active in the absence of 4-HT.

High-throughput sequencing of 11 previously described off-target sites that are modified by wild-type Cas9:sgRNA complexes targeting the EMX, VEGF, and CLTA loci revealed that both intein-Cas9 variants when treated with 4-HT for 12 h exhibit substantially improved specificity compared to that of wild-type Cas9 (Fig. 7, Tables 4, 6, and 7). On-target:off-target indel modification ratios for both intein-Cas9 variants were on average 6-fold higher, and as much as 25-fold higher, than that of wild-type Cas9 (Fig. 4b-d). In the absence of 4-HT, the genome modification specificity of both intein-Cas9 variants was on average 14-fold higher than that of wild-type Cas9 in the absence of 4-HT (Fig. 8), presumably resulting from the much lower activity of the intein-Cas9 variants in the absence of 4-HT4-6.

Since intein-Cas9s can result in slightly lower on-target modification rates compared to wild-type Cas9 (Fig. 4a), we sought to verify that the improvements in specificity among the intein-Cas9s were not simply a result of reduced activity. Both on- and off-target activity of Cas9 has been shown to be dependent on the amount of Cas9 expression plasmid transfected4-6. By transfecting lower amounts of the wild-type Cas9 expression plasmid, we compared intein-Cas9s with wild-type Cas9 under conditions that result in very similar levels of on-target modification. To minimize potential differences in transfection efficiency, we supplemented with a plasmid that does not express Cas9 so that the same total amount of plasmid DNA was transfected into each sample. High-throughput sequencing revealed that wild-type Cas9 shows slightly improved specificity, as expected, as the on-target cleavage rate is reduced. The intein-Cas9 variants, however, remain substantially more specific than wild-type Cas9 at similar on-target DNA cleavage rates (Figs. 9-11, Tables 6 and 8). For example, intein-Cas9(C574) and wild-type Cas9 (80 ng) have virtually identical on-target DNA cleavage rates (both 6.4%) at the EMX locus but all four off-target sites are modified at an average of 4-fold lower frequencies ($P < 1 \times 10^{-13}$) by intein-Cas9(C574) than by wild-type Cas9. These findings indicate that specificity improvements of intein-Cas9 variants do not simply arise from differences in overall genome editing activity.
Intein 37R3-2 can be activated by other estrogen receptor modulators. To enable intein-Cas9 applications in which endogenous β-estradiol is present, we inserted into the estrogen receptor ligand-binding domain a point mutation (G521R) that renders the domain more specific for 4-HT. This mutation slightly reduces affinity for 4-HT but almost abolishes affinity for β-estradiol. The addition of this mutation to intein-Cas9(S219) eliminates the ability of β-estradiol to trigger Cas9 activity (Fig. 12).

The intein-Cas9 variants developed here demonstrate small-molecule control of Cas9 function, thereby enhancing genome-modification specificity. The use of ligand-dependent Cas9 variants provides greater control over genomic modification efficiencies and specificities than is currently achievable with constitutively active or transcriptionally regulated genome editing. This approach can synergize with other specificity-augmenting strategies such as direct delivery of transient Cas9 protein into cells, using truncated guide RNAs, paired Cas9 nickases, or FokI-dCas9 fusions. This approach could also be applied to other genome engineering proteins to enable, for example, small-molecule control of TALE-based or Cas9-mediated transcriptional regulators.
EMX On  GAGTCCGAGCAGAAGAAGAAGGG  (SEQ ID NO: XX)
EMX Off 1  GAGgCCGAGCAGAAGAAagACGG  (SEQ ID NO: XX)
EMX Off 2  GAGTCtAGCAGgAGAAAGAGaG  (SEQ ID NO: XX)
EMX Off 3  GAGTcaAGCAGAAGAAGaAGaG  (SEQ ID NO: XX)
EMX Off 4  GAGTtaAGCAGAAGAAGAAGaG  (SEQ ID NO: XX)

VEGF On  GGGTGGGGGGAGTTTGCTCCTGG  (SEQ ID NO: XX)
VEGF Off 1  GGaTGGaGGGAGTTTGCTCCTGG  (SEQ ID NO: XX)
VEGF Off 2  GGGaGGGtGGAGTTTGCTCCTGG  (SEQ ID NO: XX)
VEGF Off 3  cGGgGGaGGGAGTTTGCTCCTGG  (SEQ ID NO: XX)
VEGF Off 4  GGGgaGGGGaAGTTTGCTCCTGG  (SEQ ID NO: XX)

CLTA On  GCAGATGTAGTGTTTCCACACGGG  (SEQ ID NO: XX)
CLTA Off 1  aCAtATGTAGTaTTTCCACACGGG  (SEQ ID NO: XX)
CLTA Off 2  cCAGATGTAGTaTTcCCACACGGG  (SEQ ID NO: XX)
CLTA Off 3  ctAGATGaAGTGcTTCCACATGG  (SEQ ID NO: XX)

Table 3. On-target and 11 known off-target substrates of Cas9:sgRNAs that target sites in EMX, VEGF, and CLTA. List of genomic on-target and off-targets sites of the EMX, VEGF, and CLTA sites are shown with mutations from the on-target sequence shown in lower case. Protospacer-adjacent motifs (PAMs) are shown underlined.
<table>
<thead>
<tr>
<th></th>
<th>- 4-HT intein-Cas9 (S219)</th>
<th>- 4-HT intein-Cas9 (C574)</th>
<th>- 4-HT wt Cas9 (500 ng)</th>
<th>+ 4-HT intein-Cas9 (S219)</th>
<th>+ 4-HT intein-Cas9 (C574)</th>
<th>+ 4-HT wt Cas9 (500 ng)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indels Total Mod. Freq</td>
<td>Indels Total Mod. Freq</td>
<td>Indels Total Mod. Freq</td>
<td>Indels Total Mod. Freq</td>
<td>Indels Total Mod. Freq</td>
<td>Indels Total Mod. Freq</td>
<td>Indels Total Mod. Freq</td>
</tr>
<tr>
<td><strong>EMX On</strong></td>
<td>1123 59967 1.87%</td>
<td>561 56700 0.99%</td>
<td>15589 72127 2.16%</td>
<td>7434 54764 13.57%</td>
<td>5209 54997 9.47%</td>
<td>9820 55972 17.54%</td>
</tr>
<tr>
<td><strong>EMX Off 1</strong></td>
<td>3 46360 0.01%</td>
<td>4 39544 0.01%</td>
<td>1143 55334 2.07%</td>
<td>185 43554 0.42%</td>
<td>116 42432 0.27%</td>
<td>1043 41387 2.52%</td>
</tr>
<tr>
<td><strong>EMX Off 2</strong></td>
<td>8 52362 0.02%</td>
<td>3 36983 0.01%</td>
<td>540 89945 0.60%</td>
<td>20 56997 0.04%</td>
<td>22 61504 0.04%</td>
<td>412 52780 0.78%</td>
</tr>
<tr>
<td><strong>EMX Off 3</strong></td>
<td>32 66472 0.05%</td>
<td>10 49582 0.02%</td>
<td>5804 83231 6.97%</td>
<td>413 53819 0.77%</td>
<td>160 56140 0.29%</td>
<td>4149 67153 6.18%</td>
</tr>
<tr>
<td><strong>EMX Off 4</strong></td>
<td>146 76633 0.19%</td>
<td>57 60976 0.09%</td>
<td>11817 86566 13.65%</td>
<td>2413 76405 3.16%</td>
<td>574 50867 1.13%</td>
<td>6561 62651 10.47%</td>
</tr>
<tr>
<td><strong>VEGF On</strong></td>
<td>359 34089 1.05%</td>
<td>379 44841 0.85%</td>
<td>3815 42732 8.93%</td>
<td>1285 35096 3.66%</td>
<td>1179 38909 3.03%</td>
<td>1120 23157 4.84%</td>
</tr>
<tr>
<td><strong>VEGF Off 1</strong></td>
<td>214 49383 0.43%</td>
<td>117 40358 0.29%</td>
<td>14578 71764 20.31%</td>
<td>2951 34729 8.50%</td>
<td>2272 48512 4.68%</td>
<td>6262 42489 14.74%</td>
</tr>
<tr>
<td><strong>VEGF Off 2</strong></td>
<td>29 34582 0.08%</td>
<td>10 21753 0.05%</td>
<td>2551 43775 5.83%</td>
<td>288 19326 1.49%</td>
<td>273 35253 0.77%</td>
<td>1199 28117 4.26%</td>
</tr>
<tr>
<td><strong>VEGF Off 3</strong></td>
<td>18 47664 0.04%</td>
<td>4 43171 0.01%</td>
<td>1743 82128 2.12%</td>
<td>167 45573 0.37%</td>
<td>107 56967 0.19%</td>
<td>679 42675 1.59%</td>
</tr>
<tr>
<td><strong>VEGF Off 4</strong></td>
<td>58 56732 0.10%</td>
<td>33 44096 0.07%</td>
<td>14114 116598 12.10%</td>
<td>1465 37619 3.89%</td>
<td>1229 88062 1.40%</td>
<td>3159 33446 9.45%</td>
</tr>
<tr>
<td><strong>CLTA On</strong></td>
<td>2087 48566 4.30%</td>
<td>930 51240 1.81%</td>
<td>16930 88447 19.14%</td>
<td>5691 39290 14.48%</td>
<td>4348 56815 7.65%</td>
<td>7974 59031 13.51%</td>
</tr>
<tr>
<td><strong>CLTA Off 1</strong></td>
<td>8 79008 0.01%</td>
<td>8 72536 0.01%</td>
<td>3361 111154 3.02%</td>
<td>286 79836 0.36%</td>
<td>32 72909 0.04%</td>
<td>1468 72166 2.03%</td>
</tr>
<tr>
<td><strong>CLTA Off 2</strong></td>
<td>3 69103 0.00%</td>
<td>0 76788 0.00%</td>
<td>75 78021 0.10%</td>
<td>0 25019 0.00%</td>
<td>11 64317 0.02%</td>
<td>18 36863 0.05%</td>
</tr>
<tr>
<td><strong>CLTA Off 3</strong></td>
<td>1 44342 0.00%</td>
<td>2 49937 0.00%</td>
<td>94 51070 0.18%</td>
<td>13 38264 0.03%</td>
<td>4 42814 0.01%</td>
<td>78 58340 0.13%</td>
</tr>
</tbody>
</table>

Table 4. Raw sequence counts and modification frequencies for data plotted in Figs. 4, 7, and 8. Total: total number of sequence counts. Modification frequency: number of indels divided by the total number of sequences listed as percentages.
Table 5. P-values for comparisons between conditions in Fig. 2a. P-values were obtained using the Fisher exact test and adjusted for multiple comparisons using the Benjamini-Hochberg Method.

<table>
<thead>
<tr>
<th></th>
<th>intein-Cas9(S219) (+ 4-HT vs. - 4-HT)</th>
<th>intein-Cas9(C574) (+ 4-HT vs. - 4-HT)</th>
<th>wt Cas9 (+ 4-HT vs. - 4-HT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMX On</td>
<td>&lt; 3.3 x 10^{-16}</td>
<td>&lt; 3.3 x 10^{-16}</td>
<td>1</td>
</tr>
<tr>
<td>VEGF On</td>
<td>&lt; 3.3 x 10^{-16}</td>
<td>&lt; 3.3 x 10^{-16}</td>
<td>1</td>
</tr>
<tr>
<td>CLTA On</td>
<td>&lt; 3.3 x 10^{-16}</td>
<td>&lt; 3.3 x 10^{-16}</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>+ 4-HT intein-Cas9 (S219)</td>
<td>+ 4-HT intein-Cas9 (C574)</td>
<td>+ 4-HT wt Cas9 (500 ng)</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------------------</td>
<td>---------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td></td>
<td>Indels</td>
<td>Total</td>
<td>Mod</td>
</tr>
<tr>
<td>EMX On</td>
<td>5446</td>
<td>66039</td>
<td>8.25%</td>
</tr>
<tr>
<td>EMX Off 1</td>
<td>134</td>
<td>65439</td>
<td>0.20%</td>
</tr>
<tr>
<td>EMX Off 2</td>
<td>26</td>
<td>69924</td>
<td>0.04%</td>
</tr>
<tr>
<td>EMX Off 3</td>
<td>438</td>
<td>81696</td>
<td>0.54%</td>
</tr>
<tr>
<td>EMX Off 4</td>
<td>1907</td>
<td>87678</td>
<td>2.18%</td>
</tr>
<tr>
<td>VEGF On</td>
<td>1633</td>
<td>51546</td>
<td>3.17%</td>
</tr>
<tr>
<td>VEGF Off 1</td>
<td>3132</td>
<td>67908</td>
<td>4.61%</td>
</tr>
<tr>
<td>VEGF Off 2</td>
<td>347</td>
<td>38567</td>
<td>0.90%</td>
</tr>
<tr>
<td>VEGF Off 3</td>
<td>84</td>
<td>52871</td>
<td>0.16%</td>
</tr>
<tr>
<td>VEGF Off 4</td>
<td>1067</td>
<td>52667</td>
<td>2.03%</td>
</tr>
<tr>
<td>CLTA On</td>
<td>4230</td>
<td>46334</td>
<td>9.13%</td>
</tr>
<tr>
<td>CLTA Off 1</td>
<td>169</td>
<td>72881</td>
<td>0.23%</td>
</tr>
<tr>
<td>CLTA Off 2</td>
<td>2</td>
<td>40883</td>
<td>0.00%</td>
</tr>
<tr>
<td>CLTA Off 3</td>
<td>5</td>
<td>45599</td>
<td>0.01%</td>
</tr>
</tbody>
</table>

Table 6. Raw sequence counts and modification frequencies for data plotted in Fig. 4b-d, and 9-11. Total: total number of sequence counts. Modification frequency: number of indels divided by the total number of sequences listed as percentages.
<table>
<thead>
<tr>
<th>Independent</th>
<th>Experiment 1</th>
<th>Independent</th>
<th>Experiment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ 4-HT intein-Cas9(S219) vs. + 4-HT wt Cas9 (500 ng)</td>
<td>+ 4-HT intein-Cas9(C574) vs. + 4-HT wt Cas9 (500 ng)</td>
<td>+ 4-HT intein-Cas9(S219) vs. + 4-HT wt Cas9 (500 ng)</td>
<td>+ 4-HT intein-Cas9(C574) vs. + 4-HT wt Cas9 (500 ng)</td>
</tr>
<tr>
<td>EMX On</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 3.9 x 10^-16</td>
</tr>
<tr>
<td>EMX Off 1</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 3.9 x 10^-16</td>
</tr>
<tr>
<td>EMX Off 2</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 3.9 x 10^-16</td>
</tr>
<tr>
<td>EMX Off 3</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 3.9 x 10^-16</td>
</tr>
<tr>
<td>EMX Off 4</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 3.9 x 10^-16</td>
</tr>
<tr>
<td>VEGF On</td>
<td>2.8 x 10^-12</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 3.9 x 10^-16</td>
</tr>
<tr>
<td>VEGF Off 1</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 3.9 x 10^-16</td>
</tr>
<tr>
<td>VEGF Off 2</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 3.9 x 10^-16</td>
</tr>
<tr>
<td>VEGF Off 3</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 3.9 x 10^-16</td>
</tr>
<tr>
<td>VEGF Off 4</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 3.9 x 10^-16</td>
</tr>
<tr>
<td>CLTA On</td>
<td>-</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 3.9 x 10^-16</td>
</tr>
<tr>
<td>CLTA Off 1</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 2.4 x 10^-16</td>
<td>&lt; 3.9 x 10^-16</td>
</tr>
<tr>
<td>CLTA Off 2</td>
<td>9.1 x 10^-5</td>
<td>4.4 x 10^-3</td>
<td>1.4 x 10^-4</td>
</tr>
<tr>
<td>CLTA Off 3</td>
<td>1.3 x 10^-7</td>
<td>1.5 x 10^-14</td>
<td>3.1 x 10^-15</td>
</tr>
</tbody>
</table>

**Table 7.** P-values for comparisons between conditions in Fig. 7. P-values were obtained using the Fisher exact test and adjusted for multiple comparisons using the Benjamini-Hochberg Method.
Table 8. P-values for comparisons between conditions in Figs. 9 and 10. All conditions were treated with 4-HT. P-values were obtained using the Fisher exact test and adjusted for multiple comparisons using the Benjamini-Hochberg Method.

<p>| Condition  | intein-Cas9(S219) vs. wt Cas9 (500 ng) | intein-Cas9(S219) vs. wt Cas9 (260 ng) | intein-Cas9(S219) vs. wt Cas9 (140 ng) | intein-Cas9(S219) vs. wt Cas9 (80 ng) | intein-Cas9(S219) vs. wt Cas9 (500 ng) | intein-Cas9(S219) vs. wt Cas9 (260 ng) | intein-Cas9(S219) vs. wt Cas9 (140 ng) | intein-Cas9(S219) vs. wt Cas9 (80 ng) | intein-Cas9(S219) vs. wt Cas9 (500 ng) | intein-Cas9(S219) vs. wt Cas9 (260 ng) | intein-Cas9(S219) vs. wt Cas9 (140 ng) | intein-Cas9(S219) vs. wt Cas9 (80 ng) | intein-Cas9(S219) vs. wt Cas9 (500 ng) |
|------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|
| EMX On     | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 1                                   | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 1                                   | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 1                                   | &lt; 3.9 x 10^{-16}                    | 0.56                                | 1                                   |
| EMX Off 1  | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 6.7 x 10^{-9}                       | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 0.84                                | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 1                                   | &lt; 3.9 x 10^{-16}                    | 3.7 x 10^{-12}                      | 3.9 x 10^{-16}                     |
| EMX Off 2  | &lt; 3.9 x 10^{-16}                    | 4.3 x 10^{-12}                     | 2.4 x 10^{-15}                     | 1.7 x 10^{-6}                       | 1                                   | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 1                                   | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 1                                   | &lt; 3.9 x 10^{-16}                    | 7.1 x 10^{-2}                       | 3.9 x 10^{-16}                     |
| EMX Off 3  | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 2.0 x 10^{-6}                       | 1                                   | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 1                                   | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 1                                   | &lt; 3.9 x 10^{-16}                    | 7.5 x 10^{-13}                      | 3.9 x 10^{-16}                     |
| EMX Off 4  | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 1                                   | 1                                   | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 1                                   | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 1                                   | &lt; 3.9 x 10^{-16}                    | 1.0 x 10^{-11}                      | 3.9 x 10^{-16}                     |
| VEGF On    | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 1.7 x 10^{-6}                      | 1                                   | 1                                   | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 1                                   | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 1.9 x 10^{-6}                      | 1                                   | 1                                   |
| VEGF Off 1 | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 3.8 x 10^{-12}                      | 1                                   | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 1                                   | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 3.9 x 10^{-16}                      | 9.6 x 10^{-2}                      | 3.9 x 10^{-16}                     |
| VEGF Off 2 | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 5.7 x 10^{-2}                       | 1                                   | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 1                                   | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 3.9 x 10^{-16}                      | 3.9 x 10^{-16}                     | 3.9 x 10^{-16}                     |
| VEGF Off 3 | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 5.9 x 10^{-13}                      | 1.8 x 10^{-6}                       | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 1                                   | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 3.9 x 10^{-16}                      | 3.9 x 10^{-16}                     | 3.9 x 10^{-16}                     |
| VEGF Off 4 | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 1                                   | 1                                   | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 1                                   | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 9.9 x 10^{-11}                     | 9.9 x 10^{-11}                     | 9.9 x 10^{-11}                     |
| CLTA On    | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 1                                   | 1                                   | 1                                   | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 1                                   | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 3.9 x 10^{-16}                      | 3.9 x 10^{-16}                     | 3.9 x 10^{-16}                     |
| CLTA Off 1 | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 0.21                                | 0.74                                | 1                                   | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 1                                   | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | &lt; 3.9 x 10^{-16}                    | 2.0 x 10^{-9}                      | 3.9 x 10^{-16}                     | 2.0 x 10^{-9}                      |
| CLTA Off 2 | 1.4 x 10^{-4}                      | 1.6 x 10^{-4}                      | 0.13                                | 9.4 x 10^{-2}                       | 0.23                                | 4.6 x 10^{-6}                       | 5.4 x 10^{-6}                       | 4.8 x 10^{-2}                       | 3.5 x 10^{-2}                      | 5.7 x 10^{-2}                       | 5.7 x 10^{-2}                       | 0.11                                | 0.11                                | 0.11                                |
| CLTA Off 3 | 3.1 x 10^{-15}                     | 2.3 x 10^{-2}                      | 0.12                                | 0.11                                | 0.29                                | 3.5 x 10^{-15}                     | 9.3 x 10^{-3}                       | 5.7 x 10^{-2}                       | 5.7 x 10^{-2}                      | 5.7 x 10^{-2}                       | 5.7 x 10^{-2}                       | 0.16                                | 0.16                                | 0.16                                |</p>
<table>
<thead>
<tr>
<th></th>
<th>Indels</th>
<th>Total</th>
<th>Modification frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EMX On</strong></td>
<td>8</td>
<td>78943</td>
<td>0.01%</td>
</tr>
<tr>
<td><strong>EMX Off 1</strong></td>
<td>1</td>
<td>42232</td>
<td>0.00%</td>
</tr>
<tr>
<td><strong>EMX Off 2</strong></td>
<td>4</td>
<td>79008</td>
<td>0.01%</td>
</tr>
<tr>
<td><strong>EMX Off 3</strong></td>
<td>60</td>
<td>113629</td>
<td>0.05%</td>
</tr>
<tr>
<td><strong>EMX Off 4</strong></td>
<td>5</td>
<td>104159</td>
<td>0.00%</td>
</tr>
<tr>
<td><strong>VEGF On</strong></td>
<td>0</td>
<td>60667</td>
<td>0.00%</td>
</tr>
<tr>
<td><strong>VEGF Off 1</strong></td>
<td>2</td>
<td>111409</td>
<td>0.00%</td>
</tr>
<tr>
<td><strong>VEGF Off 2</strong></td>
<td>0</td>
<td>52048</td>
<td>0.00%</td>
</tr>
<tr>
<td><strong>VEGF Off 3</strong></td>
<td>4</td>
<td>88105</td>
<td>0.00%</td>
</tr>
<tr>
<td><strong>VEGF Off 4</strong></td>
<td>2</td>
<td>123559</td>
<td>0.00%</td>
</tr>
<tr>
<td><strong>CLTA On</strong></td>
<td>491</td>
<td>68600</td>
<td>0.72%</td>
</tr>
<tr>
<td><strong>CLTA Off 1</strong></td>
<td>10</td>
<td>116033</td>
<td>0.01%</td>
</tr>
<tr>
<td><strong>CLTA Off 2</strong></td>
<td>6</td>
<td>75723</td>
<td>0.01%</td>
</tr>
<tr>
<td><strong>CLTA Off 3</strong></td>
<td>4</td>
<td>53885</td>
<td>0.01%</td>
</tr>
</tbody>
</table>

**Table 9.** Raw sequence counts and modification frequencies (for cells transfected with wild-type Cas9 (500 ng) but without a targeting sgRNA, in the presence of 4-HT). Total: total number of sequence counts. Modification frequency: number of indels divided by the total number of sequences listed as percentages.
Sequences
[00151]

Intein 37R32:

TGCCTTGCCGAGGGTACCCGAATCTTCGATCCGGTCACTGGTACAACGCATCGCA
TCGAGGATGTTGTCGATGGGCGCAAGCCTATTCATGTCGTGGCTGCTGCCAAGGA
CGGAACGCTGCTCGCGCGGCCCGTGGTGTCCTGGTTCGACCAGGGAACGCGGGA
TGTGATCGGGTTGCGGATCGCCGGTGGCGCCATCGTGTGGGCGACACCCGATCAC
AAGGTGCTGACAGAGTACGGCTGGCGTGCCGCCGGGGAACTCCGCAAGGGAGAC
AGGGTGGCCGGACCGGGTGGTTCTGGTAACAGCCTGGCCTTGTCCCTGACGGCCG
ACCAGATGGTCAGTGCCTTGTTGGATGCTGAGCCCCCCATACTCTATTCCGAGTA
TGATCCTACCAGTCCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACCTG
GCAGACAGGGAGCTGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAGGCTTT
GTGGATTTGACCCTCCATGATCAGGCCCACCTTCTAGAACGTGCCTGGCTAGAGA
TCCTGATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCAGGGAAGCTACTGTT
TGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAGGGCATGGT
GGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCGCATGATGAATCTG
CAGGGAGAGGAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATTCTGGAGTGT
ACACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCG
AGCCCTGGACAAGATCACGGACACTTTGATCCACCTGATGGCCAAGGCAGGCCT
GACCCTGCAGCAGCAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCAC
ATCAGGCACATGAGTAACAAAGGAATGGAGCATCTGTACAGCATGAAGTACAAG
AACGTGGTGCCCCTCTATGACCTGCTGCTGGAGATGCTGGACGCCCACCGCCTAC
ATGCGGGTGGTTCTGGTGCTAGCCGCGTGCAGGCGTTCGCGGATGCCCTGGATGA
CAAATTCCTGCACGACATGCTGGCGGAAGGACTCCGCTATTCCGTGATCCGAGAA
GTGCTGCCAACGCGGCGGGCACGAACGTTCGACCTCGAGGTCGAGGAACTGCAC
ACCCTCGTCGCCGAAGGGGTTGTCGTGCACAACTGC (SEQ ID NO: XX)
[00152]

Cas9-NLS-3xFLAG:

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQR
TFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFA
WMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE


[00153] Intein-Cas9(C80)-NLS-3xFLAG:

MDKKYSIGLIDGINTSVGAVITDEYKVPSSHKFKVLGNTDHRHIKKNLIGALLFDSGE
TAEATLKRKTARRRTYTRRNRICLAEGRIFRDPTVGTTHRREDVDGPKIHPHVAAA
KDGTLLARPVSVSFDQGTRDVIHLRIAGGAIYAVATPDHKVLTTEYGWRAAGELKRG
DRVAPGPGSSGLSALS LTADQMVSALLDAEPPLYSEYDPTSPFSEASMMGLLTNLALR
DRELVMINAWKRPVGFVDTLHIDQAHLLEARLEILMGVWRSMEHPGKLFFA
POLLDDRNQGKCEGMVGEFLMALTSSFRMNNLQEEFVECLVSILLNSGVYTFLS
STLKLSEKIHRLSKDTIITDTHLMAKGALTLOOHORLALQLLILSHIRHSMNGK
MEHLYMSKYKNVVPYDLLELDMAHRHLHAGGGSARVOAFADALDDKFLHDMIL
AEGRLYIREVLPTRRARTFDELVEELHTVLEVGVHCNOEIFSNEMAKVDSS
FFHRLEESFLVEEDKKHERPHFNVIDVEAYHEKPYTIYRRKLKLVDSTKADRLRL
YLAHMIKHFKRGHFLGLPNDVSDKFIQVLQVTYQLFEENPASGKVDAAIL
SLARKSRRLENLIQLPGKEKKNGFLGFLIALSGLTPNFKSNFLLADAKQLSKDT
YDDDDLNLLAQQIAKLYDLDSDLLDVLNEITLKPALPSASMKRYDE
HQLDITLCLAKVRQOLPEKYEIIFFDQQSNKGYAYIDGGAQSEEFYFKFPKLEMD
GEELVVLNREDLRKQRTDFNSPSQIPHIHLGELHAILRQDEFYFPLKDNREKI
ELTRIPYYVGVDELPVMGRHPENIVEMARENQTTQKQGNRMRKRIEIEGK
ELGSQILKEHPVENTQLQNEXKLYLLYQNQRDMYVDQELDINLSDYDVHIPQSVS
LKDSDINKLTRSDNRGKSDNVPEEVEKMKNYWQLNACLITQRKDFTNLT
AEGRLSLEDKAIFIKRQVETRQTIKVHVAPIQRDMKNTYDNEKLERVKTIL
KLYSDFKDFQFYKVRINNYYHAAHYANVGTALIKKYPLESEFYGYDVKVY
DVRKMIASAQKYFFSNIMNFKEKITEIAGEIRKPLJETNGEIIIVWD
KGDRFATVRKLSMPQVIVKKTETVTQGFSKESILPKRNSDKLIARKKDWDPPKY
GGFDSTPVAVYLVVAKVEKGKKSLKSVKELIGIMERSSEFKPIDFLEAKGYKE
VKKDLIIKLPYSLFELENGRKRLASAGELQKNELALPSKVYNFLASHYELKL
GSPEDNEQKQFLVEQHKLHDEIEQISEFSKRVILADNLKVLASAYNKHRDKPIRE
QAENIIHFLTITNLGAPAAFKYFDTTIDDRKRYSTKTEVLDAIHLHSITGLYTERIDLSQ
LGGDGSPKKKRKCVSSDYKHDGDYKHDIDYKDDDDKAAG (SEQ ID NO: XX)

[00154] Intein-Cas9(A127)-NLS-3xFLAG:

MDKKYSIGLIDGINTSVGAVITDEYKVPSSHKFKVLGNTDHRHIKKNLIGALLFDSGE
TAEATLKRKTARRRTYTRRNRICLAEGRIFRDPTVGTTHRREDVDGPKIHPHVAAA
KDGTLLARPVSVSFDQGTRDVIHLRIAGGAIYAVATPDHKVLTTEYGWRAAGELKRG
DRVAPGPGSSGLSALS LTADQMVSALLDAEPPLYSEYDPTSPFSEASMMGLLTNLALR
DRELVMINAWKRPVGFVDTLHIDQAHLLEARLEILMGVWRSMEHPGKLFFA
POLLDDRNQGKCEGMVGEFLMALTSSFRMNNLQEEFVECLVSILLNSGVYTFLS
STLKLSEKIHRLSKDTIITDTHLMAKGALTLOOHORLALQLLILSHIRHSMNGK
MEHLYMSKYKNVVPYDLLELDMAHRHLHAGGGSARVOAFADALDDKFLHDMIL
AEGRLYIREVLPTRRARTFDELVEELHTVLEVGVHCNOEIFSNEMAKVDSS
FFHRLEESFLVEEDKKHERPHFNVIDVEAYHEKPYTIYRRKLKLVDSTKADRLRL
YLAHMIKHFKRGHFLGLPNDVSDKFIQVLQVTYQLFEENPASGKVDAAIL
SLARKSRRLENLIQLPGKEKKNGFLGFLIALSGLTPNFKSNFLLADAKQLSKDT
YDDDDLNLLAQQIAKLYDLDSDLLDVLNEITLKPALPSASMKRYDE
HQLDITLCLAKVRQOLPEKYEIIFFDQQSNKGYAYIDGGAQSEEFYFKFPKLEMD
GEELVVLNREDLRKQRTDFNSPSQIPHIHLGELHAILRQDEFYFPLKDNREKI
ELTRIPYYVGVDELPVMGRHPENIVEMARENQTTQKQGNRMRKRIEIEGK
ELGSQILKEHPVENTQLQNEXKLYLLYQNQRDMYVDQELDINLSDYDVHIPQSVS
LKDSDINKLTRSDNRGKSDNVPEEVEKMKNYWQLNACLITQRKDFTNLT
AEGRLSLEDKAIFIKRQVETRQTIKVHVAPIQRDMKNTYDNEKLERVKTIL
KLYSDFKDFQFYKVRINNYYHAAHYANVGTALIKKYPLESEFYGYDVKVY
DVRKMIASAQKYFFSNIMNFKEKITEIAGEIRKPLJETNGEIIIVWD
KGDRFATVRKLSMPQVIVKKTETVTQGFSKESILPKRNSDKLIARKKDWDPPKY
GGFDSTPVAEYSLVAVKEVKGKKSLKSVKELIGIMERSSEFKPIDFLEAKGYKE
VKKDLIIKLPYSLFELENGRKRLASAGELQKNELALPSKVYNFLASHYELKL
GSPEDNEQKQFLVEQHKLHDEIEQISEFSKRVILADNLKVLASAYNKHRDKPIRE
QAENIIHFLTITNLGAPAAFKYFDTTIDDRKRYSTKTEVLDAIHLHSITGLYTERIDLSQ
LGGDGSPKKKRKCVSSDYKHDGDYKHDIDYKDDDDKAAG (SEQ ID NO: XX)
DNA sequence:

```
NVVPLYDLLEMLDAHRLHAGGSGASRVOAFADALDDKFLHDMALAEGLRYSVIRE_VLPTRRARTFDLEVEELHTLVAEGVYVHCYHEKYTIYHLRKLVDSTDKADLRLY
YLALAHMIKFRGHFLIEGDLNPDNSVDEVKFLIQFLVQTYNQLFEENPISGVDAAKAIL
SARLKSRRLENLIAQLPGEKKNFLGNGLIALSLGLTPNFKSNDLAEADKLQLSKDT
YDADDLNNLQAIGDQYADLFLAANLSDAISDILRNVKTEITAKPALSAMIKRYYDEH
HQDLTLALKVRQOLPEKYKIEFDQSQKNGYAGIDDGASQFYEFKPIKMDGTE
TEELVKLNRLDRKRTFDNNSHIPQHHGLHELHAILLRQEDFYFPFLKIRLNEKEIKI
TFRIPYYVGPlARGNSFAWMTRKSEITTPRNFEVEVDGKASAQSFIERTMNDK
LPNEKVLHSSLLEYFTVYNELTKYVYVTEGMPKAPLSGEQKGKAIADLDLFKNTKR
VTVKQKLDLYFLKQIECFDESIEGVDERFSNLGTKYDYHHLKAIKDKDLFDNENEDILE
DIVTLTLFEDREMIERLKTAYAHLFDKVMQLKRRRTYGWRLSRRLINGIRDSKQ
SGKTLISFLSDGFFANRNNMLHIHDLDFEKLIDQIKAQVSGQGDLHEHIANLAGSPA
IIKGGilQTVKVDHELKVGMHRKPHENIVIERAMNRTTOQKQKNSREMCRKIEEGIK
ELGSQILKEHPVENTOLNEKLTYLYQLNQRMDYVDQELDRNLSDYDVIDHVPSQSF
LKDDSIDNKLVTORSDKNRGKSNDVPSSEEVKMKMYWQRLLNKLITQKRFDNSLTK
AERGGLSELDKAGFIRKLQVELTRQITKHQAQILDSRNTKDYDENDKLIKEKVKL
KSLSDFKRDKFQFYKVEINNYHHADNYLANVGGTALIKYKPLESEVFYGDVKYV
DVRKMIASEQIEQIKATYYFYSSNIMNNFFKTEITLNGEIRKLPIETRNGGETIEWD
KGRDFATVRKLMSQPVPNIVKKTENVQTGFGKESILKPNRSKDLKILAKKWDQDDK
GGFDSPTVAYAVVAKEGKSSKLKSVKELLGITIMERSFEFKPDIFLAEKGYKE
VKKDDLILPKYSLEFENFERMKRLASELQGKQNLALPSKYVNFYLSHLYEKLK
GSPDNEQILQFVEHQHLYDEIEQISEFKVIRADNLKVLASYNHKRDPIRE
QAENIIHFLTTLNLGAPAFFYDFTDNRKRTSTKEVDLATHIQSTGGLYETRIILSQ
GLGGDSPPKKKVRKVSDDYKDHGDYKDHDDYKDDDAAK (SEQ ID NO: XX)
```

Reference:

[Intein-Cas9(T146)-NLS-3xFLAG: MDKKYSGILGDTNSVGVAVITDEYVKPSKKFKVLGNTDHSIKKNNLIGALLFDSGE TAEATRLKRTTRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRRLEESFLVEEDKKE RHPIFGNIVDVEAYHEKYTIYHLRKLVDSCLAEGTRIFDPVTGPHTTRHIDVVDGRK PIHVAAAKDGTLLARRPVSWSFQGTRVDVGRIAGAIVWATPDVKLTEYGWRA AGELRKGDRVAGPGGSGNSLASLTLADQMVMALLDEPPILYSEYDPTSPFSEASMM GLUTNADRELVHWNMMVIAKDWHERAWLEILMIQLWVSMKPGKLLFAIPNLLDDRNOGKCGCEGMVEIFDMLLATSSBFRRMMNLLQGEEFVCLKSIIILNSGVGTTFSLSTLLEKEDDIHELALDKDIITDLHMAKGLTLOOHORLAQLLLISI RHMSNKGMEHLYSMKNNVPYILDDLMEMLDAHRLHAGGSGASRVOAFADALDD KFLHDMALAEGLRYSVIRE_VLPTRRARTFDLEVEELHTLVAEGVYVHCYHEKYTIYHLRKLVDSTDKADLRLY

[00155] Intein-Cas9(T146)-NLS-3xFLAG:
AERGGLSELDKAGFIKRQLVETRQITKHVAQILDMSNMTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVRINNNYHHADAYLNAVGTALIKKYPKLESEFYGDYKVY
DVRKMIAKSEQUEIGKATAKYFFYSNIMFFKTEITLANGEIRKPLIETNGETGEIVWD
KGRDFATVRMPQVQTFQGSGKESILPKNDSKLIAIKKWDWPKKY
GGFDSPTVAVSMLLVVAKVEGKSKKLSVKVEKLLEGITIMERSSFEKNPIDFLEAKGYKE
VKDLDIIKLPKYSLELGKKMLASAGELQKGNELALPSKVYNFLASHYELKL
GSPEDNEQKQLFVEQHKHYLDEIEQISEFSKRVLADANLKDVLASYNKHRKDPIRE
QAEIIHFLTLSNLGAAPAFKYDFTTIDRKRYSTKEVLATLHIQSITGLYETRIDLQS
LGGDGSPKKKRVSSDYKHDGDYKHDIDYKDDDDKAAG (SEQ ID NO: XX)

[00156] Intein-Cas9(S219)-NLS-3xFLAG:
MDKKYSIGLDGTNSVGWAVITDEYKVPSKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKTARRRTRKRNRICLQEIIFSNEMAKDIVDSFFHLEESFLVEEDKKE
RHPFIGNVDEVAYHEKYTIHYLRKVLVDSTDKALRILYVLALAHMIKEFRGHFLEIG
DLNPDSVDKLFIOLVQYLOFENPIASVDKAILSLKSCLAEGTRIFDPY
TGTTTRIEMDVGPIHVAAAKDGTLLARPSVSWDODTRVIGRIAGGAIVW
ATPDHKVLTEYGRAWAAGLKRKGDRVAGPGGGSNSLALSADOMYSALLDAAPELL
YSEYDPTSPFSEASMGMILNTNLADRELVHINWAVKPVPFDLTDHQAHILLERA
WLEIMILVGWRSMHEPGKLLFAPNNLDRNOOGKCVEGMVEIFDMMLLATSSRFRMM
NQEGEEFVLCIKILLNGVYTFLSSTLKLSEKDHIMRALKDTITMLMAKAGTLT
OOCHARLOALLILLSHIRHMSNKGMHLYSMKYKNVPLYDDLLEMLDHARLHAG
GGSASVRVOAFADDLKFLMLEGRYIVRELTRLTPRATFDLEEHTLVAE
GVVHNCRRLNLIAOLPGKEKNKLGLNLLALSGLGTPFNKSNFDLAEADKLQSLK
TYEEDDDDLNLLAQIDQYADFLAILAKNIALDSLIDLVRNTEITKAPLSASMIKRYDE
HHQDLTLLKALVRQQLPEKYKEIFFDFQNKSGNYAGYGAGSQQEYFIFKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHIQLHGLHAILRRQEDFYPFLDNRKIEK
LTFRIPYYGVLARGNRFAWMTKRKSETTPWNFEVVDKGASAQSFIERMTNDK
NLPNKVLKPHSLLEYFTVYNELKTVKYVEGMRKPAFLSGLGEQKKAIVDLLFKTNR
KVTVDQKLDYFKKIECFDSVIEGVEDRNLSAGTYHDLLKIIKDKFLDNEENEDI
LEDIVLLTLFEDREMIEELRKYAHFLFDKVMQKLRRRTGWWLRSKCLINGRID
KQSGKTTLDLFLSDKGFANRFMQHLHDSLTFKEDIQKAQPSGQGDSLHEHIANLAG
SPAIIKGIGLQTVDKVMVRKHPENIEMARENQQTQGKNSRERMKRIE
GIKELSLQKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRSLDYVDHIVP
QSFKLKDSINDKVLCNKSNDRGKNSVNPSEVVKMKNYRQLLAIRKTPRFDN
LTKAERGLSELDKAGFIKRQLVETRQITKHVAQILDMSNMTKYDENDKLIREVKVIT
LKSKLVDFFKQFQYKVRINNNYHHADAYLNAVGTALIKKYPKLESEFYGDY
KVDVYKRMIAKSEQUEIGKATAKYFFYSNIMFFKTEITLANGEIRKPLIETNGETGEI
VWDKGRDFATVRKVLMPQVQNFVCKTEVQQTGGSKESILPKNDSKLIAIKKWDW
PKKYGGFDSPTVAVSMLLVVAKVEGKSKKLSVKVEKLLEGITIMERSSFEKNPIDFLEAK
GYKEVKKDLIIKLPKYSLELGKKMLASAGELQKGNELALPSKVYNFLASHYELKL
GSPEDNEQKQLFVEQHKHYLDEIEQISEFSKRVLADANLKDVLASYNKHRKDPIRE
QAEIIHFLTLSNLGAAPAFKYDFTTIDRKRYSTKEVLATLHIQSITGLYETRIDLQS
LGGDGSPKKKRVSSDYKHDGDYKHDIDYKDDDDKAAG (SEQ ID NO: XX)

[00157] Intein-Cas9(T333)-NLS-3xFLAG:
MDKKYSIGLDGTNSVGWAVITDEYKVPSKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKTARRRTRKRNRICLQEIIFSNEMAKDIVDSFFHLEESFLVEEDKKE
RHPFIGNVDEVAYHEKYTIHYLRKVLVDSTDKALRILYVLALAHMIKEFRGHFLEIG
DLNPDSVDKLFIOLVQYLOFENPIASVDKAILSLKSCLAEGTRIFDPY
TGTTTRIEMDVGPIHVAAAKDGTLLARPSVSWDODTRVIGRIAGGAIVW
ATPDHKVLTEYGRAWAAGLKRKGDRVAGPGGGSNSLALSADOMYSALLDAAPELL
YSEYDPTSPFSEASMGMILNTNLADRELVHINWAVKPVPFDLTDHQAHILLERA
WLEIMILVGWRSMHEPGKLLFAPNNLDRNOOGKCVEGMVEIFDMMLLATSSRFRMM
NQEGEEFVLCIKILLNGVYTFLSSTLKLSEKDHIMRALKDTITMLMAKAGTLT
OOCHARLOALLILLSHIRHMSNKGMHLYSMKYKNVPLYDDLLEMLDHARLHAG
GGSASVRVOAFADDLKFLMLEGRYIVRELTRLTPRATFDLEEHTLVAE
GVVHNCRRLNLIAOLPGKEKNKLGLNLLALSGLGTPFNKSNFDLAEADKLQSLK
TYEEDDDDLNLLAQIDQYADFLAILAKNIALDSLIDLVRNTEITKAPLSASMIKRYDE
HHQDLTLLKALVRQQLPEKYKEIFFDFQNKSGNYAGYGAGSQQEYFIFKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHIQLHGLHAILRRQEDFYPFLDNRKIEK
LTFRIPYYGVLARGNRFAWMTKRKSETTPWNFEVVDKGASAQSFIERMTNDK
NLPNKVLKPHSLLEYFTVYNELKTVKYVEGMRKPAFLSGLGEQKKAIVDLLFKTNR
KVTVDQKLDYFKKIECFDSVIEGVEDRNLSAGTYHDLLKIIKDKFLDNEENEDI
LEDIVLLTLFEDREMIEELRKYAHFLFDKVMQKLRRRTGWWLRSKCLINGRID
KQSGKTTLDLFLSDKGFANRFMQHLHDSLTFKEDIQKAQPSGQGDSLHEHIANLAG
SPAIIKGIGLQTVDKVMVRKHPENIEMARENQQTQGKNSRERMKRIE
GIKELSLQKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRSLDYVDHIVP
QSFKLKDSINDKVLCNKSNDRGKNSVNPSEVVKMKNYRQLLAIRKTPRFDN
LTKAERGLSELDKAGFIKRQLVETRQITKHVAQILDMSNMTKYDENDKLIREVKVIT
LKSKLVDFFKQFQYKVRINNNYHHADAYLNAVGTALIKKYPKLESEFYGDY
KVDVYKRMIAKSEQUEIGKATAKYFFYSNIMFFKTEITLANGEIRKPLIETNGETGEI
VWDKGRDFATVRKVLMPQVQNFVCKTEVQQTGGSKESILPKNDSKLIAIKKWDW
PKKYGGFDSPTVAVSMLLVVAKVEGKSKKLSVKVEKLLEGITIMERSSFEKNPIDFLEAK
GYKEVKKDLIIKLPKYSLELGKKMLASAGELQKGNELALPSKVYNFLASHYELKL
GSPEDNEQKQLFVEQHKHYLDEIEQISEFSKRVLADANLKDVLASYNKHRKDPIRE
QAEIIHFLTLSNLGAAPAFKYDFTTIDRKRYSTKEVLATLHIQSITGLYETRIDLQS
LGGDGSPKKKRVSSDYKHDGDYKHDIDYKDDDDKAAG (SEQ ID NO: XX)
IVLTLTFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDSTLFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILTQVVDVLKVVMGRHKPENIVIEMARENQTQKQKNSRERMKRIEEGIK
ELGSQILKEHPVENTQLQNEKYLYYLQNGRDMDYVDQELDRSDLVDVHIVPQSF
LKDSDIDNKLVRTRSDKNRGKDSNVPEEVVKKMKNYWRQLLNAKLITQRKFDNLTK
AERGGLSEDLKAGFIKQVLTRQITKHVAQLDSRMNTkyDENDKLIREFIXVVTILKS
KLVSDFRDFFQYKVEINNYHHAHYDLYNAVGTLAIKPKLESEFYVDKYVY
DVKRMIAKSEQIGAKATAKYFYSNIMFFKTEITLANEIRKRPLIETNGETIEIVWD
KGRDFATVRKLVMQPVNIVKKTVEQTQGFSEIKSRPNSDKLIAKRDWDPPKYY
GGFDSPTVAYSVLAVKVEKGGSSKLKVIELGKIGITISERSSSFEKPNIDFLAEAGYKE
VKDLIIKLPYSLEFENKRMASAGELQKGNELALPSKIVYNFLYLASHYEKLK
GSPDENEQQLFVEKHYDLEHEIQISEFSKVRILADANKLVAYNKHRDKPIREF
QAENIIHLFTLNLGAAPAFKTYDFTTDIRKRYSTKEVLDAIHLQ5ITGLYETRIDSQ
LGGDGSPKKKKRVSSDYKHDGHDGKYHDIDIDKYDDDDKAAG (SEQ ID NO: XX)
[00159] Intein-Cas9(C574)-NLS-3xFLAG:
MDKKYSIGLTDVGWAVITDEYKVPSSKKFKVLNGTDRHSSIKKNLALLFDGSE
TAAEATRLKTARRYVRNGKRYLQEISMEAKVVDSSFHRLSESSLVEODEKKE
RHPIFGNVDVAYHEKPUHYLKLVLSTDKADLPILYALALAHIMKFRGGFLIEG
DLPNDSVDKLQYVTQNQLENFBNIPANSVGDAKAILSAKLSSRLRENLIAPG
GEKKGNLGFLNGSALSLGTNPQFNDSLALQEDALQLSDKYDDDDDHAQQ
DLFIAAKNLSDEALLSLRVTENITKAPSASMISMKYDEHHDQLTLLKALVRQOLPE
KYKEIFQDSKNGYAGIDGGASQEEFYKFKPILEKMDGTEELVKNREDLRLLKRQ
TFDNGSIPHQHILGELHAILRQDQFYFPLKDNREIEKILTFRIPYVYPLARGNSRFA
WMTRKSEETrpPWNEEVEVDGKASAQSFIERMTNDLNPENKVLPKHSLYETFV
YNELTJKVYVTGMRKPAFLSGBEOKIAVDLLFKNTRKVTQQLKEDYFKKIECLA
EGTRIFDPVTTHIREDVVDGKRPIHVAAAKDTLLARPVVSWFDQGTRDVIGLRI
AGGAIWVATPHDKVLEYGWRAGELRKGDRVARPGGNSLALSLTADQVMASL
LDAEPPILHSEYDPTSPSEASMMLNLADXDELHVMINWAARVPGFDVLTLHDAQ
HLLERAWLEILMIGLVRMSMEHPGKLFAFNLNDRNOGKCEVMGEIFDMLLATS
SRFRMMNQLGEEVFVCLSDKIIINNSGYVTFLSSSTLKSLEEKDHIALDKITDTLHIHMA
KAGLTLUUOHORLAAQQLLSSLHHRMNSGKMHELYSMKYKNVPLYDLLEMLDA
HLRHAGGSGARVAOFADALDDKFLHDLAMEGLRYSVIREVLPTRARTFDLEEE
LHTLVAGVHVVINCDSLVEISGVEDRFNASLGTYHDLLKIIKDKDFNEDENEILED
IVLTLLTFLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDSTLFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILTQVVDVLKVVMGRHKPENIVIEMARENQTQKQKNSRERMKRIEEGIK
ELGSQILKEHPVENTQLQNEKYLYYLQNGRDMDYVDQELDRSDLVDVHIVPQSF
LKDSDIDNKLVRTRSDKNRGKDSNVPEEVVKKMKNYWRQLLNAKLITQRKFDNLTK
AERGGLSEDLKAGFIKQVLTRQITKHVAQLDSRMNTkyDENDKLIREFIXVVTILKS
KLVSDFRDFFQYKVEINNYHHAHYDLYNAVGTLAIKPKLESEFYVDKYVY
DVKRMIAKSEQIGAKATAKYFYSNIMFFKTEITLANEIRKRPLIETNGETIEIVWD
KGRDFATVRKLVMQPVNIVKKTVEQTQGFSEIKSRPNSDKLIAKRDWDPPKYY
GGFDSPTVAYSVLAVKVEKGGSSKLKVIELGKIGITISERSSSFEKPNIDFLAEAGYKE
VKDLIIKLPYSLEFENKRMASAGELQKGNELALPSKIVYNFLYLASHYEKLK
GSPDENEQQLFVEKHYDLEHEIQISEFSKVRILADANKLVAYNKHRDKPIREF
QAENIIHLFTLNLGAAPAFKFYDFTTDIRKRYSTKEVLDAIHLQ5ITGLYETRIDSQ
LGGDGSPKKKKRVSSDYKHDGHDGKYHDIDIDKYDDDDKAAG (SEQ ID NO: XX)
Intein-Cas9(T622)-NLS-3xFLAG:

MDKKEYS1GDLGNTSVGWAYTIDYKEVKPSKKKFKVGLGNTDHISSIKKLNIGALFDGS
tAEATRLKRTARRARRYTRRRKCN1RCLQIEFSNEMAVKDSFHRLEESFLVEEDKKHE
HRPIFPGGNVDEVAYHEKYPIYHRLKLVSTDKADLRILYLaAHIKFKGRFHFLIEG
DLNPNDSDVDKLFIQLVQNYLQFEENPINASVDAKAILSRKSSRKLLENLIALQP
GEKKNGLFGNLALSGLTPFNKSNFDLAEDAKLQLSKDSTDLDLDLLAQIIGDYA
DLFLLAANKLSDLALSIDLRLVNEITKAPLSASMIRYDEHHQDLTLLKALVRQQLPE
KYKEIIEFDQSKNGYAGYIDGGASQEEFYFKIKPLEKMDGTEELLVKLENREDLKRQR
TDFN5PSIFQHSLGHELHAIIRQEDDFYPFLDNRKEIKLTRIPYPVGLPARGNSRFA
WMTRKSEETrPWNNFEVVDKGAASQFIERMTNDFDKNLPKVLPKSHLLLEYFTV
YNELTKVYATEMGKPAFLSQEKKAIVDLLFKTNKTRVKVQLKEDYFKKIECFD
SVEISVGEDRFNASLGLYHDLKKIYKDKFLDNEENEDILEDIVLCLAEIGTRIDFDPVTG
TTHRIEVDVGRKPHPIHVAAA acquitted all the arguments

Intein-Cas9(S701)-NLS-3xFLAG:

MDKKEYS1GDLGNTSVGWAYTIDYKEVKPSKKKFKVGLGNTDHISSIKKLNIGALFDGS
tAEATRLKRTARRARRYTRRRKCN1RCLQIEFSNEMAVKDSFHRLEESFLVEEDKKHE
HRPIFPGGNVDEVAYHEKYPIYHRLKLVSTDKADLRILYLaAHIKFKGRFHFLIEG
DLNPNDSDVDKLFIQLVQNYLQFEENPINASVDAKAILSRKSSRKLLENLIALQP
GEKKNGLFGNLALSGLTPFNKSNFDLAEDAKLQLSKDSTDLDLDLLAQIIGDYA
DLFLLAANKLSDLALSIDLRLVNEITKAPLSASMIRYDEHHQDLTLLKALVRQQLPE
KYKEIIEFDQSKNGYAGYIDGGASQEEFYFKIKPLEKMDGTEELLVKLENREDLKRQR
TDFN5PSIFQHSLGHELHAIIRQEDDFYPFLDNRKEIKLTRIPYPVGLPARGNSRFA
WMTRKSEETrPWNNFEVVDKGAASQFIERMTNDFDKNLPKVLPKSHLLLEYFTV
YNELTKVYATEMGKPAFLSQEKKAIVDLLFKTNKTRVKVQLKEDYFKKIECFD
SVEISVGEDRFNASLGLYHDLKKIYKDKFLDNEENEDILEDIVLCLAEIGTRIDFDPVTG
TTHRIEVDVGRKPHPIHVAAA acquitted all the arguments

[00160] Intein-Cas9(T622)-NLS-3xFLAG:

[00161] Intein-Cas9(S701)-NLS-3xFLAG:

[227]
LSLTADQMVSALLDAEPPILYSEYDPTSPFSEASMMGLLTNLADRELVHMINWAKRV
PGFVDLTLHDQAHLLERAWLEILMGLVWRSMEHPGKLFAPNLLLDRNOGKCVEG
MVEIFDMLLATSSFRMNQLGEEFVCLKSIILLNSGVYTFLSTLKLSEEDHIHRA
LDKITDTLLHMAKGTLOOHORAILLQLLSHRHMNSKNGMEHLSMYKYNV
PLYLDILLEMELDAHRHAGGSGASRVOAFADALDDKFLHDMLEAGLRSYVIREVLPT
RRARTFDLEVEELHTLVAEGVVHVNNCLTFKEDIIOKACQVSOGDSLSEHEIANLAGSPA
IKKGILQTVKVVGRHPPENIEVAMRANNQTQKQKNSERMKRIEEGIK
ELGSQILEHPVENTQLNEKLYLYQLONGRMDYVDQEELDINLDYVDHIVPQSF
LKDSDINDKVLTRSDKNRGSKDNSVPEEVVKMMKNYWQRLLNKLITQRFDNLTK
AERGGLSELDAGKFIRKLQVTRQITKHAVQLDSRMNTKYDENDKLIBREVKVTILKS
KLVSDFRKDFQFYKVRINNHYHAIYALVNAVTLAKKYPLESEFYVGYDVKY
RVKRIAKSEEQIGKATAKYYFVSNIMNFETTLEIGEIRKPLEIETGIEVW
KGDFATVRLVSMPQVNIKKTVETQGGFSESLPKRNSDKLIAKKDKWPDKPY
GGFDSTVSAYVLAGSVEGKLKLKLKSLKVSKVLEKILGMRMSEFKNPDLEAKGYKE
VKKDLIIKLPKYSLFELENKRKRMLASAGELQKNGELALPSKYVNFLASHYELK
LGSPEDEKQKLFVEQHKHYLDIEIQIFSEKRFVLADANLDKLVLSAYKHRDKPIRE
QAEINIIHLFTLTNLGAPAAFKFYDFTTIDRRKRTYSTKVDLATHIQSITGGYETDLRS
LGGDGSPKKRRVYDDKHDYGKYDHDIDYKDDDDKAAG (SEQ ID NO: XX)

[00162] Intein-Cas9(AD728)-NLS-3xFLAG:
MDKKYSIGLDIGTNSGVGAVITDEYKPSKFKVLGNTDRSHIKKNLIGALLLDFSGE
TAEATRLKRTARRYTRRRKNRICTQEIFSNEMAKVDSSFHRLEESFLVEEDKKE
RHPIFGNVEDAYHEKYPTIYHLKLKVVDSTKDRLYLIALAHMIKFRGHFLIEG
DLNPDSVDKLFIQQLVQTYNLFEENPINASVADAKAISLARKNSRLLENLIAQP
GEKKGFLNGLSILGTPNKFSDLAEDAKLQLKSDYDNDLNAIQDQYACL
DLAFLAAKNLSDAILSDILRNTVEITKLPSASMIMKYDEHHQDTLTLLCALVRQQLPE
KYEIEEFDQSKNGYAGYIDGGSQEEFYKFIKPILEKMDGTEELLKLKINREDLKQR
TFDNSHPHQHILGELHAILHRQQEDFYPFLKDRENIEKILTRFPYYVPLAGNRSFA
WMTRKSEEITITPNFEVENKDGSAQSFIERMNTDFKLPNEKVLPHSLLEYTYFT
YNELTJKVYTEGMRKPAFLSGLQEKQIKAVIILDFTKRNVTQKVLKDKEYFKKIECFD
SVEISGVEDRFNASLGHTYHDLLKIKDKDFDLNEDILEDIVLTLTFFEDREIERL
KTYAHLFDDKVMQLKRRYTGWLGRSLINGIRDQSKGTLDFKSLDGFANNN
FMOLIHDDSLFKEDIOOQACQVSOGDSLSEHEIANLCLAEGRIFDPTGVVTHRIDEY
DGRKPIHVAAAKDGGTAPVSVWFDQGTRVGLRIAGIAVWATPPKVLTEY
GWRAAGELRKGDRVAGPGGSGNLASLTAQMSVALLDAEPPILEYDPTSPFSE
ASMMGLLTNLADRELVHMINWAKRVPGVDFDLTLHDQAHLLERAWLEILMGLVW
RSMHEPMGKLFAPNLLLDRNOGKCVEG
MLIILLNSGVYTFLSTLKLSEEDHIHRALDKITDTLLHMAKGTLOOHORAILLQLL
ILSHIRHMSKNGMEHLSMYKYNVPLYLDILLEMELDAHRHAGGSGASRVOAFAD
ALDDKFLHDMLEAGLRSYVIREVLPTRRARTFDLEVEELHTLVAEGVVHVNNCGSPI
AKKGIQLTQKVVDHPHKKRPVNIEMARENTQQTQKQKNSERMKRIEEGIK
ELGSQILEHPVENTQLNEKLYLYQLONGRMDYVDQEELDINLDYVDHIVPQSF
LKDSDINDKVLTRSDKNRGSKDNSVPEEVVKMMKNYWQRLLNKLITQRFDNLTK
AERGGLSELDAGKFIRKLQVTRQITKHAVQLDSRMNTKYDENDKLIBREVKVTILKS
KLVSDFRKDFQFYKVRINNHYHAIYALVNAVTLAKKYPLESEFYVGYDVKY
RVKRIAKSEEQIGKATAKYYFVSNIMNFETTLEIGEIRKPLEIETGIEVW
KGDFATVRLVSMPQVNIKKTVETQGGFSESLPKRNSDKLIAKKDKWPDKPY
GGFDSTVSAYVLAGSVEGKLKLKLKSLKVSKVLEKILGMRMSEFKNPDLEAKGYKE
VKKDLIIKLPKYSLFELENKRKRMLASAGELQKNGELALPSKYVNFLASHYELK
LGSPEDEKQKLFVEQHKHYLDIEIQIFSEKRFVLADANLDKLVLSAYKHRDKPIRE
Intein-Cas9(T995)-NLS-3xFLAG:

MDKKYSIGLDI GTNVGWAVITDE YKVPSSKKFKVGLNTRGLHSDIKKLNIGALLFDSGE TAEATRLKRTA T RRYTRRNIR CYLQIEIFSNEMAKVDSFHFHLEESFLVEEDKKHE RHPIFGNIVDEVAYHEKPTYIHYLRLKCLVDSTDKADLRLIYLALAHMIKFRGHFLIEG DLNPDNSVDKLFIQVLQYQLFENIPINASGVDAKIALSARLSKRRLENLIAQLP GEKKNGLFGNIALSLGLGTPNFKSNFNLADAKLLQLSDKTYDDDDLNLLAQIDGQYA DLFLAANLSDAILSDILVRNTEITKAPLSASMKRYDEHHQDLTLKLVRQQLPE KYKEIFFDQSKNGAYIDGGAQSEEEFYKFIKPLEKMDGTEELVKLNRLELRKQR TFDNSGIPHQIHLGELHAILRRQDQEDFYPF PKDNAKREIEKILTFRIPYYVGPLARDSRFA WMTRKSEETntrPNFEEVVDKGASAAQFIERMTNGTDNLNPEKLPKSSLLEYFTV YNELTKVKYVTEGMRKPAFLSEQKKAI VDLFPTKRNKTVKQLKEDYF KKIECFD SVEISVGVDRENSNLDQHDLLKIDKDFNLEEDILEVTLTLFEDREMIE LKTYAHLFDDKV MKQLRKYTWGRLSLRKIN GIRDQKSQKTILD LSDKDFANN FMQLIHI DSSLTFKEDIQKAQVSGQGDHEHIANL GASPAIKGQILTQVKTVDVVMG RHPKENPSQTDQKGGKNKSRMRRIEKIGELQSILKEHPVENTQL QNEKLYLYLQNRD VQELINDLSXYDVHQPVFLK DDSIDNKLVT SDSR NRGKDNVPEE VVKMKY WQRLNSSL TKRQDNLTKRQDG EELQGRLSELDKAGF IQ RQLVETRQTIKHV AQILD LSMNTKYDENKLIREKVTLKSLVSDFKDFQFYKV REINNYHAHADLYNAVGCALAEGRDFPVTGTVH DQRKHQHVVAA KDTLLARPVS WFDQTR VDI GGAIAVWD PAHKGVTE GRLAEGELR GDR VAGPGGSNSLALS LTAQM VSDAALD EEPROMILSE YDTPSFPEASM GMGTNLAD LR ELVHMNSAKVRGPVFLTLHDQAHLLERWAMLGVILWSMHEPKLFLA PNL LLDRRNOGK VEGMVEI FDMULATSSFRMNMLQGEVFVLKSIILLNSGYVF TSSTL KSLLEKDIHRA LDKIDTTLHI LMAKAGTLLOOHORL AQLLLIL SHRHMSNKGME HLYSMK YKVYAD VLTDLEMLA HDRLHAGGS GVSROAFL DDDKF HLM D A GLRSYVIREVP TRRTADFLEEL HTLVAEGV VHNCALIK KYPK LSEFVYGYDV KVDERMIKAQ SLEQIEGKATAKYFQSNIMNKF TEILTLANG EIERKRPLIETNETGEI VWDKGRDFATVRKVLSMPQVNVIKK VTEVTTGGSKESILPKRN DKLARIK DWDPP KYGGFDSPVTAPYVSAL VAVK EGGKSS LKV EGMITGERSSEFKPIDFLEA GKYVEKDKLIKLPSLYSLEFENGRKMLASALGK OKG NEALSPPKLVYNFLYSLHY EKLKGSPEDEQK VFQHIKYLDEIEQISEFVRVILADANL DKLVSAYNKH DK PIREQAENIIHLFTLNLGAPA AFK YFDTIDRKRKYSTKEVLDALTLDHQSITGLEYTRI DLSQ LGGDGSPKKKRKVSSDYKDHYDGYKDDDDKAAG (SEQ ID NO: XX)

Intein-Cas9(S1006)-NLS-3xFLAG:

MDKKYSIGLDI GTNVGWAVITDE YKVPSSKKFKVGLNTRGLHSDIKKLNIGALLFDSGE TAEATRLKRTA T RRYTRRNIR CYLQIEIFSNEMAKVDSFHFHLEESFLVEEDKKHE RHPIFGNIVDEVAYHEKPTYIHYLRLKCLVDSTDKADLRLIYLALAHMIKFRGHFLIEG DLNPDNSVDKLFIQVLQYQLFENIPINASGVDAKIALSARLSKRRLENLIAQLP GEKKNGLFGNIALSLGLGTPNFKSNFNLADAKLLQLSDKTYDDDDLNLLAQIDGQYA DLFLAANLSDAILSDILVRNTEITKAPLSASMKRYDEHHQDLTLKLVRQQLPE KYKEIFFDQSKNGAYIDGGAQSEEEFYKFIKPLEKMDGTEELVKLNRLELRKQR TFDNSGIPHQIHLGELHAILRRQDQEDFYPF PKDNAKREIEKILTFRIPYYVGPLARDSRFA WMTRKSEETntrPNFEEVVDKGASAAQFIERMTNGTDNLNPEKLPKSSLLEYFTV YNELTKVKYVTEGMRKPAFLSEQKKAI VDLFPTKRNKTVKQLKEDYF KKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDKVKMQVQLKRRRTYGWRLSRKLINGIRDQSQTGKTLFLKSDGFANRN
FMQLIHDDSLTFKEDIQVASQGQGDSHELHIANLASHPAIKKGILQTVKVVDVELKV
VMGRHKPVENIMARENQTQKQNRSRMKRIIEGKELIQSILKEHVENTQL
QNEKLYLYLQNQRDMYVDQELDINRSLDDYDHDVHPQFLKSDDSNKLVTDRSK
NRGKSNDNVPEEVSQKMKNYWRQLNLNNLTIQRFKDFNLTKAERGLSELDAQFGIK
RQLQVETRQTIKHVAENKRMTNDKLERIKVTLKLSSDFKDFQFYKV
REINNYHHHAYLNAAVGTLIKKYPKLE CLAEITRIFDPVTGTTHRIEVDVGRK
PIHVAAKAQDGTLARRPVSSFDQGTRDVIGLRAGIINAVATPDKHVLTLEYGWRA
AGELRKGRDVAPGGSGNLSLTLTADQVMVSALLDAEPPILYSEYDTPSFSEASMM
GLTMLADRELVHMINWAKRVPGFVLDTLDQAIHLKRLARWLEILMIGLWWRSMEH
PGKLQANLDRNOGKCLVQGPVEMGDFMILLATSSRRMNQLQGEVFCKLSSLII
SNGYTFSLSTLKSLEEKDHIHALDLKIDTLTLHMAKAGLTLQOOHORLAAQLLILLSI
RHMNSKGHELMGKVNVPVLYLTEKLSLILDEAHRLHAGSSGASVOAFADLADD
KFLHDLMAEGLRYSVIREVLPTRRARTFDLEVLEELHTLVAGVVHVHCEFVIYGDYK
VYVYRKMIAKSEQEIGKATAKYFFYSNMIFKTEITLANEIRKRPIEETNGETEIVG
WDKGRDFATVRKLSPMQVNVKKTVEQTGFSKESIFPNKSRNDLJARKDWDPK
YKGFDSPVAYSVVLKAYAVYKVGGSKKLSKSVKELGGIERTIMESSFENPFDLAEAKG
YVEKVKDLJKLPYSLEENGRKMLASAGEQKGNELALPSKYVNVFLAYLASHYKE
KLKGSPEDNEQKQVHEQHLYDEIEITSEQEFSKRVILADALNLKLYSAYNKHRRDKP
IREQAENIIHFLTLNLGAPAFAFYDFTDIDCRKYSTKELVTDLSLHQSsGTYETRIDL
LSQLGGDGSPKIKRVSSDYKDHGDYDKHDIDYKDDKADDA

[00165] Intein-Cas9(SL 154)-NLS-3xFLAG:
MDKKYSIGLDTNSVGWAVITDEYKVPKSKFKLVGLNTRHSSIKNLIGALLFDSGE
TAEATRLKRTARRRRYTTRRNRCYLAGQFESNEMAKVDSSFHRLLESFLVVEEKKHE
RHPIGNDVEAYHEKYPIYHLKRLVVDSTKDALRLLYLALAHMIKFRGFLILLID
DLVPDNSDVFKLFLFQVLYQUNYNNLEENPISASGVDAKILSRKSLDEIIAPQP
GEKKGNLFGNIALSLTGNPSFDNAELAKQLSKDSTDILNLLAQIGDQYADLF
FLAAKNLSDAILSDLIRNTEITDKAPLSAMIKYRDEHHQDILLLKALVQQLPE
KYKEIFDDQSKNYAGYIDGASQEEFYFKIFKPILEKMDGTEELVKLNRDLRRKQR
TFDNSIPHIQLHELHAIIRRRQEFYPFPLDKNREIKIFTRIPYPVGLPAGRSFA
WMTKRKSeqTrPWNFEEVVDGKASAQSFIERMTFDKLNPEKVLPHSLLFYFTV
YNELTKVYVTEMGAPFALSEQKAVIDLLEFKTRKVTQVQLKEDYFKKIECFD
SVEISGVEDRFNASLGYTHQDLKIDDFKDEENEDILEDIVLTLTLDREDIMHEL
KTYAHLFDKVKMQVQLKRRRTYGWRLSRKLINGIRDQSQTGKTLFLKSDGFANRN
FMQLIHDDSLTFKEDIQVASQGQGDSHELHIANLASHPAIKKGILQTVKVVDVELKV
VMGRHKPVENIMARENQTQKQNRSRMKRIIEGKELIQSILKEHVENTQL
QNEKLYLYLQNQRDMYVDQELDINRSLDDYDHDVHPQFLKSDDSNKLVTDRSK
NRGKSNDNVPEEVSQKMKNYWRQLNLNNLTIQRFKDFNLTKAERGLSELDAQFGIK
RQLQVETRQTIKHVAENKRMTNDKLERIKVTLKLSSDFKDFQFYKV
REINNYHHHAYLNAAVGTLIKKYPKLE CLAEITRIFDPVTGTTHRIEVDVGRK
PIHVAAKAQDGTLARRPVSSFDQGTRDVIGLRAGIINAVATPDKHVLTLEYGWRA
AGELRKGRDVAPGGSGNLSLTLTADQVMVSALLDAEPPILYSEYDTPSFSEASMM
GLTMLADRELVHMINWAKRVPGFVLDTLDQAIHLKRLARWLEILMIGLWWRSMEH
PGKLQANLDRNOGKCLVQGPVEMGDFMILLATSSRRMNQLQGEVFCKLSSLII
SNGYTFSLSTLKSLEEKDHIHALDLKIDTLTLHMAKAGLTLQOOHORLAAQLLILLSI
RHMNSKGHELMGKVNVPVLYLTEKLSLILDEAHRLHAGSSGASVOAFADLADD
KFLHDLMAEGLRYSVIREVLPTRRARTFDLEVLEELHTLVAGVVHVHCEFVIYGDYK
VYVYRKMIAKSEQEIGKATAKYFFYSNMIFKTEITLANEIRKRPIEETNGETEIVG
WDKGRDFATVRKLSPMQVNVKKTVEQTGFSKESIFPNKSRNDLJARKDWDPK
YKGFDSPVAYSVVLKAYAVYKVGGSKKLSKSVKELGGIERTIMESSFENPFDLAEAKG
YVEKVKDLJKLPYSLEENGRKMLASAGEQKGNELALPSKYVNVFLAYLASHYKE
KLKGSPEDNEQKQVHEQHLYDEIEITSEQEFSKRVILADALNLKLYSAYNKHRRDKP
IREQAENIIHFLTLNLGAPAFAFYDFTDIDCRKYSTKELVTDLSLHQSsGTYETRIDL
LSQLGGDGSPKIKRVSSDYKDHGDYDKHDIDYKDDKADDA

(SEQ NO: XX)
PLYDLLLEMLDAHRLHAGGSGRVOAFADALDDKFLHDMLAEGLRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVHVNCKKKLKSVKELLGITIMERSSSFEKNPFDLAEKGYKEVKKDLIKKYSLEFENGKRMALASAGELQGKNEALPSKYVFNFLYLASHYEKLKGSPEDNEQFKHVYKDEPQIEFSEKVRVILADANLKVLSAYNKHDKPIREQAENIIHLFTLNLGAAPAACKFYDTITDRKRYSTKEVLDATLHIQSSrGLEYTRIDLSQLGGDGSPKKKKRKVSSDYKDHGDGYKDHIDYKDDDDKAAG (SEQ ID NO: XX)

[00166] **Intein-Cas9** (S1159) - **NLS-3xFLAG:**
MDKKYSIGLDGTSVGAIVTEYKVPSKKFKVGLNTDRHSLIKKLGLIGALLFDGSETAETRLKRTARRYTRRRKNRICYQLIEIFNEMAVKDDSFFHREELVESFLVEEDEKHKTIPIFNVEDVAYHEKYTIYLKRKVLSTDSDKADLLYIYLAHMIKFRGFLIEGDLNPDNSVDKLFLIQVLTQNYLQMFAPNIAVGDAKSLARLKSRRLENLIAQLPGEKKNLGFNGNLALSLGTPFQSMDALQEDAKLQSKDCTYDQDLLNLQAGIDQYADLFMLAKNLSDAILSLIDILVNTIEITKAPSAMIKYREDEHQLTDLLKALVRQQLPEKYKIEFDFQOSKGNYAGIDGASQEFYKFKIPEKMDGTEELVVLKNEDLLRKQR TFDNSPIQHLGFHELIALIEDPFNLKDNREEIKILFTTRPYYGVLARGSFAWMTRKSEETrrPWNFEVEVDKGAASQSFIERMNTFDKLNPEKVLPSLLLEYFTTVYNELTKVYVETGMRKPAFLSEQQKAIAVDLPFKTNRKVTKVQKLEDYFKKIECFDSVEISGVEDRFNLASYNKLDKDLESMNEDILEVDILITLTEFDREIEERRKYLAHLFDVKMDQKLRKYTGWLRSKLINGIRDKQSGKTLDFLSDKGAFANNFMQLIHDSDLTFKEDIQKAVQGQGDQGSLHHEIANLASPAGKIGLQTVDKVLVMGRHKPENIVIEMARENQTPTQKQKSNRSERMKRIIEEIGKELGSGQLKHEPVENTQINEKLQLYLQNRMDQILSDYDHDVPSQLFKDDSDNKLVTTRSDKNRGKDNVPSEEVVKKMNYYWQLNLAKLITQRFDNLTAERGLSLEDKAGFIQRQLVETRQITKAVQILDMSRMTKDYENDKLREVQVTKLKSVLTDVRKDFQYKYRKRENNYHHAHDAYLNAVNGTALIIKYYKPELESEFYVGYDYVRKMIAKSEQEIQGKATAKYFFYSNIMNFKTIEILANGERIRKPLIETNGGETEIQVWDKGRDFATVRKVLSMPQVNIKKETVQTGFSKESILPKRNSSDKLARKKDWDPPKYYGGFSPTVAYSLLVLVAKKEKGSKKLKCLAEETRIPFVGTTTIRIEDTVDGRKPHVVAADGTILLARPVVSFWFDQGTRDVGILRIAGQJAVWATPDHKVLTLEYGWAAGELKRDVARGPSGSNLALSMTAQSMVSAALLDAEPPPLYSEYDPTSPFEASMMGGLLTNLADRHELVMINWAKRKPFGVDROLTLDHAHLERAWLEILMIGLVRWSMMEHPKGLFAPINNLDRONOGCVGMEVEIFDMLMATTSRFRMNLIQGEFVCLKsvilleSILLNSGYTVLFSLKLSEKDHIHARDKIDTDLHMAKKAGLTOOOHORNPLLLLIHSHRMSNGKHMELSYKYNVPLLDDLEMLDAHRLHAGGSGASROFAFADALDDKFLHDMLAEGLRYSVIREVLPTRARTFDLEVEELHTLVAEGVVVHVNCKKKLKSVKELLGITIMERSSSFEKNPFDLAEKGYKEVKKDLIKKYSLEFENGKRMALASAGELQGKNEALPSKYVFNFLYLASHYEKLKGSPEDNEQFKHVYKDEPQIEFSEKVRVILADANLKVLSAYNKHDKPIREQAENIIHLFTLNLGAAPAACKFYDTITDRKRYSTKEVLDATLHIQSSrGLEYTRIDLSQLGGDGSPKKKKRKVSSDYKDHGDGYKDHIDYKDDDDKAAG (SEQ ID NO: XX)

[00167] **Intein-Cas9** (S1274) - **NLS-3xFLAG:**

MDKKYSIGLDGTSVGAIVTEYKVPSKKFKVGLNTDRHSLIKKLGLIGALLFDGSETAETRLKRTARRYTRRRKNRICYQLIEIFNEMAVKDDSFFHREELVESFLVEEDEKHKTIPIFNVEDVAYHEKYTIYLKRKVLSTDSDKADLLYIYLAHMIKFRGFLIEGDLNPDNSVDKLFLIQVLTQNYLQMFAPNIAVGDAKSLARLKSRRLENLIAQLPGEKKNLGFNGNLALSLGTPFQSMDALQEDAKLQSKDCTYDQDLLNLQAGIDQYADLFMLAKNLSDAILSLIDILVNTIEITKAPSAMIKYREDEHQLTDLLKALVRQQLPEKYKIEFDFQOSKGNYAGIDGASQEFYKFKIPEKMDGTEELVVLKNEDLLRKQR
TFDNSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFA
WMTRKSEETrpWNFEEVVDKGASAASQFIERMTNFDKNLPEVKVLKHSLLYEYFTV
YNELTKVYTYEGRMRFPAFLSGEQKKAIVDLLFLKNTVRKTVQKLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLKIIKDKFDLNENNEDILEDIVTLTFREDMIEERL
KTYAHLFDKVMKQLKRRTYGTGWRSLRKLINGIRDKQSGBKTDLFLKSDGFANRN
FMQLIHDSSLTFKEDIQKAAVPSQQGDSHELHIANLPSAIPIAGLQGILQTVKV
VMGHRKPINIENQTQKQGKNSRERMKRIEIGEKESQLKHEPVNTQQLNEKLYL
YNQDMDYVQDLELNRDSYDDVDHIQPQFLKDSIDNKLVTRSDK
NRGSKDNSEEVVKMKNYWQRLNALKITQRFDNLTKERGLSLEDKAFIK
RQLVETRQTIKHVAQIRLUMTNTKYLREDKLIRVEKVIKLKVSDRFKDFQFYKV
REINOYHNHAHDYLNAVGVTALIKKYPLESEVFYDYDGKVYDVRKMIAKSEQEIG
ATAKYFFYNSNMIFKTEITLANEIRKPLIETNGETIEIVDKGRDFATVRKVLSM
PQVNIKKTEVQTGGFSILPELIRKCIAYKSRDFPKYYGDPSPTVAYSVLV
AVKEKGSKKKLKSKVEILGITISSERTSSEFKNPILDGEYKVYDKLIKLKPYL
LENGRKRMLASAGELKQGNEALPSKYVNYFLASHYELKYGSPEDNQKQFOEQ
HKYHLDEIEIQICLAEGRTRDPVGTGTHRIDEVDVRKPIHVVAAKDGTLLARPV
SWFDQGTRDVLIGRIAGGAIWATPDHWLEYGWAAGELKGRDVAAPGGSNG
SLASLTADQMSVALLSEYDPTSPFSEASSMMGLLTNADRELVHMINWAKRVP
GFDVLTLHDQAHLLERAWLEILMLGLVWMSHEPGKLFAPLLLDNRNOGKC
VEGMEVEIDFLDMALTSSRFMNLQFCAEVCLKISSLHYYFTLSLKSLEEKDH
HRLDKIDTTLHIHMAKGLTLOOOHORLAIQLLILSHIRHMSNKMEHLYSMKYK
NVPLYLDLLEMLDAHLRHAGSASROAFAADALDDKFLHMDLAEGLRYSVIRE
VLPTTRARTFVKLVDELADGAPGQSGLASLTADQMSVALLSEYDPTSPFSEAS
MMGLLTNADRELVHMINWAKRVPGFVVLTLHDQAHLLERAWLEILMLGLVWMSH
EPGKLFAPLLLDNRNOGKCVEGMEVFDMALTSSRFMNLQGEFVCLKISSLHYYFTL
SLKSLEEKDHRLDKIDTTLHIHMAKGLTLOOOHORLAIQLLILSHIRHMSNKMEH
LYSMKYKNNPYVLYLDLLEMLDAHLRHAGSASROAFAADALDDKFLHMDLAEGL
RYSVIREVLPTTRARTFVKLVDELADGAPGQSGLASLTADQMSVALLSEYDPT
SPFSEASMMGLLTNADRELVHMINWAKRVPGFVVLTLHDQAHLLERAWLEILMLG
LVWMSHEPGKLFAPLLLDNRNOGKCVEGMEVFDMALTSSRFMNLQGEFVCL
KISSLHYYFTLSLKSLEEKDHRLDKIDTTLHIHMAKGLTL0OOHORLAAOII
LILSHIRHMSNKJRMELHYLMSKYNVPLYLDELLEMLDAHLRHAGS
GASRVOAFADALDDKFLHMDLAEGLRYSVIREVLPTTRARTFDEEVLHTLV
GEVVHNCRLLENIQLPQEGKKNLFGNIALSLGGTLPNHFKSNDFLEADKQLSK
KDTYDDDNLQAIQDQYADFALAKNLSDAILDLRVNTEIKAPLSAMIKRYDE
HHQDLTLKALVRQPLKEYIFFDFQSKNGAYIDGASQEFEYKFIPKILEM
DGETELVKLNREDLRRKQRTFDPISHPHQILGHELHAIIRROQDFYPFKLDRNEKIEI
LTFRIPYYVGPLARGNSRFAWMTTRKSEETrpWNFEEVVDKGASAASQFIERMTNF
DKNLPEVKVLKHSLLYEYFTVYNETLTKVYTYEGRMRFPAFLSGEQKKAIVDLLFLK
TKVTQKLEDVFYFKKICFDSISVEISGVDFNASLGTYHDLKIIKDKFDLNENNED
LEDIVTLTFREDMIEERLKTAYHLFDKVMQLKRRTYGTGWRSLRKLINGIR
KQSGKTLTDFKSDGAFANRFMQLIHDSSLTFKEDIQKAAVPSQQGDSHELHIANL
GSPAIIKGLQTVKVVDLKVGMGRHKPINIENQTQKQGKNSRERMKRIE
Indel Calling Algorithm

Read 1:

@M00265:68:000000000-AA85W: 1:1101: 14923: 1642 1:N:0: 1
TGCACTTCATGACTGCTTTTGAGAAAAACACCATTAGAAAGAG_TAGATGGTT
GGGTAGTGCTCTCTCTGTACTCTTGACTCTTGACTCTATGAAATAAGGGCTCTA
TTTCAAGGGCGTGTGGTGTAAGCAAGACACATTTGAGAAAGGCT

3>A?FFFFFFGGGGGHHHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGH

Read 2:

@M00265:68:000000000-AA85W: 1:1101: 14923: 1642 2:N:0: 1
CTCACCGGCGGAAGAAGGTAACCTATTTCTTCTGAGCCTCTTCTCTCATTGCTTT
CAACCCATCAGCCTTGTGCAAAATAGAG_CCCTTATTCATAGTAGACAA_GAGTCT
AAGCAGAAGAGAGAGCCACTACCCAACCACCATCTCTACTCTATGAGT

3>AACFFBBBGGGGGHHHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGHGHGGH

Step 1: Search for sequences (or reverse complements) flanking the on/off target sites in both Illumina reads from the following set:

<table>
<thead>
<tr>
<th>target site</th>
<th>5’ flanking sequence</th>
<th>3’ flanking sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMX_On</td>
<td>GAGTCGAGAGGCAAGAAGAAGGAGG</td>
<td>AGCTGGAGGGAGAAGGGGCCT</td>
</tr>
<tr>
<td>EMX_Off1</td>
<td>GAGCCGAGCAGAAGAAAGAAGG</td>
<td>CCCCTCCCTCTGCAAAATGAG</td>
</tr>
<tr>
<td>EMX_Off2</td>
<td>GAGTCCTAGGAGGAGAAGAAGG</td>
<td>GGGCAGGGGGCGACAGTACC</td>
</tr>
<tr>
<td>EMX_Off3</td>
<td>GAGTCAGAGCAGAAGAAAGAAGG</td>
<td>CCTTATTCTAGTACAGAAA</td>
</tr>
<tr>
<td>EMX_Off4</td>
<td>GAGTTAGGAGAGCAGAAGAAGAAGG</td>
<td>CATGGCAAGAGAGATGTCA</td>
</tr>
</tbody>
</table>
VEGF_On GGGTGGGGGGAGTTTGCTCCTGG GGGAATGGGCTTTGGAAAGG ACCCCCTATTTCTGACCTCC
VEGF_Of f 1 GGATGGAGGGAGTTTGCTCCTGG CATCTAAGGACGGATTTGTG
VEGF_Of f 2 GGGAGGGTGGAGTTTGCTCCTGG CATGAGAGAGGCTCCCATCA
VEGF_Of f 3 CGGGGGAGGGAGTTTGCTCCTGG CTGAGAGAGGCTCCCATCA
CLTA_On GCAGATGTAGTGTTTCCACAGGG CTGAGTAGGATTAAGATATT
CLTA_Of f 1 ACATATGTAGTATTTCCACAGGG GTTGGGAAGAGATGCATACA
CLTA_Of f 2 CCAGATGTAGTATTTCCACAGGG GCCTCCTTGATTGAGGTGTC
CLTA_Of f 3 CTAGATGAAGTGCTTCCACATGG CTCATCTAGAGTTCTTTCCA

Step 2: Extract the sequence between the target sites in both reads and ensure that it is identical (reverse complementary) in read 1 and read 2 and all positions within read 1 and read 2 have a quality score \( \geq 30 \) (Phred score \( \geq 30 \))

In above reads, CTCTTCTGCTTAGACTC is reverse complement of GAGTCTAAGCAGAAGAG

Step 3: Align extracted sequence to the reference sequence for the relevant on/off target sequence

GAGTCTAAGCAGAAGAG reference sequence
GAGTCTAAGC---------AGAAGAG sequence read

Step 4: For deletions, count only if deletion occurred in close proximity to expected cleavage site (within 8 bp of 3' end of reference sequence)

Methods and Materials

[00169] Cas9, intein-Cas9, and sgRNA expression plasmids. A plasmid encoding the human codon-optimized Streptococcus pyogenes Cas9 nuclease with an NLS and 3xFLAG tag (Addgene plasmid 43861)\(^5\) was used as the wild-type Cas9 expression plasmid. Intein 37R3-2 was subcloned at the described positions into the wild-type Cas9 expression plasmid using USER (NEB M5505) cloning. sgRNA expression plasmids used in this study have been described previously\(^{11}\). Plasmid constructs generated in this work will be deposited with Addgene.

[00170] Modification of genomic GFP. HEK293-GFP stable cells (GenTarget), which constitutively express Emerald GFP, served as the reporter cell line. Cells were maintained in "full serum media": Dulbecco’s Modified Eagle’s Media plus GlutaMax (Life
Technologies) with 10% (vol/vol) FBS and penicillin/streptomycin (lx, Amresco). 5 x 10^4 cells were plated on 48-well collagen-coated Biocoat plates (Becton Dickinson). 16-18 h after plating, cells were transfected with Lipofectamine 2000 (Life Technologies) according to the manufacturer's protocol. Briefly, 1.5 µl, of Lipofectamine 2000 was used to transfect 650 ng of total plasmid: 500 ng Cas9 expression plasmid, 125 ng sgRNA expression plasmid, and 25 ng near-infrared iRFP670 expressing plasmid (Addgene plasmid 45457). 12 h after transfection, the media was replaced with full serum media, with or without 4-HT (1 µM, Sigma-Aldrich T176). The media was replaced again 3-4 days after transfection. Five days after transfection, cells were trypsinized and resuspended in full serum media and analyzed on a C6 flow cytometer (Accuri) with a 488-nm laser excitation and 520-nm filter with a 20-nm band pass. Transfections and flow cytometry measurements were performed in triplicate.

**High-throughput DNA sequencing of genome modifications.** HEK293-GFP stable cells were transfected with plasmids expressing Cas9 (500 ng) and sgRNA (125 ng) as described above. For treatments in which a reduced amount of wild-type Cas9 expression plasmid was transfected, pUC19 plasmid was used to bring the total amount of plasmid to 500 ng. 4-HT (1 µM final), where appropriate, was added during transfection. 12 h after transfection, the media was replaced with full serum media without 4-HT. Genomic DNA was isolated and pooled from three biological replicates 60 h after transfection using a previously reported protocol with a DNAdvance Kit (Agencourt). 150 ng or 200 ng of genomic DNA was used as a template to amplify by PCR the on-target and off-target genomic sites with flanking HTS primer pairs described previously. PCR products were purified using RapidTips (Diffinity Genomics) and quantified using the PicoGreen dsDNA Assay Kit (Invitrogen). Purified DNA was PCR amplified with primers containing sequencing adaptors, purified with the MinElute PCR Purification Kit (Qiagen) and AMPure XP PCR Purification (Agencourt). Samples were sequenced on a MiSeq high-throughput DNA sequencer (Illumina), and sequencing data was analyzed as described previously.

**Western blot analysis of intein splicing.** HEK293-GFP stable cells were transfected with 500 ng Cas9 expression plasmid and 125 ng sgRNA expression plasmid. 12 h after transfection, the media was replaced with full serum media, with or without 4-HT (1 µM). Cells were lysed and pooled from three technical replicates 4, 8, 12, or 24 h after 4-HT treatment. Samples were run on a Bolt 4-12% Bis-Tris gel (Life Technologies). An anti-FLAG antibody (Sigma-Aldrich F1804) and an anti-mouse 800CW IRDye (LI-COR) were used to visualize the gel on an Odyssey IR imager.
[00173] **Statistical analysis.** Statistical tests were performed as described in the figure captions. All p-values were calculated with the R software package. p-values for the Fisher exact test were calculated using the fisher.test function, with a one-sided alternative hypothesis (alternative = "greater" or alternative = "less", as appropriate). Upper bounds on p-values that are close to zero were determined manually. The Benjamini-Hochberg adjustment was performed using the R function p.adjust (method = "fdr").

[00174] **Sensitivity limit of off-target cleavage assays.** We used paired end sequencing to identify indels caused by genomic on- and off-target cleavage. Given that published studies (see the reference below) have shown that the Illumina platform has an indel rate that is several orders of magnitude lower than the -0.1% substitution error rate, and our requirement that all called indels occur in both paired reads, the sensitivity of the high-throughput sequencing method for detecting genomic off-target cleavage in our study is limited by the amount genomic DNA (gDNA) input into the PCR amplification of each genomic target site. A 1 ng sample of human gDNA represents only -330 unique genomes, and thus only -330 unique copies of each genomic site are present. PCR amplification for each genomic target was performed on a total of 150 ng or 200 ng of input gDNA, which provides amplicons derived from at most 50,000 or 65,000 unique gDNA copies, respectively. Therefore, the high-throughput sequencing assay cannot detect rare genome modification events that occur at a frequency of less than approximately 1 in 50,000 (0.002%). When comparing between two conditions, such as wt Cas9 vs. intein-Cas9, this threshold becomes approximately 10 in 50,000 (0.02%) when using the Fisher exact test and a conservative multiple comparison correction (Bonferroni with 14 samples). See also Minoche, A. E., Dohm, J. C., & Himmelbauer, H. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. *Genome Biology* 12, R2 (2011).

**REFERENCES**


All publications, patents, patent applications, publication, and database entries (e.g., sequence database entries) mentioned herein, e.g., in the Background, Summary, Detailed Description, Examples, and References sections, are incorporated by reference in their entirety as if each individual publication, patent, patent application, publication, and database entry was specifically and individually incorporated herein by reference. In case of conflict, the present application, including any definitions herein, will control.

**EQUIVALENTS AND SCOPE**

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. The scope of the present invention is not intended to be limited to the above description, but rather is as set forth in the appended claims.

In the claims articles such as "a," "an," and "the" may mean one or more than one unless indicated to the contrary or otherwise evident from the context. Claims or descriptions that include "or" between one or more members of a group are considered satisfied if one, more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process unless indicated to the contrary or otherwise evident from the context. The invention includes embodiments in which exactly one member of the group is present in, employed in, or otherwise relevant to a given product or process. The invention also includes embodiments in which more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process.

Furthermore, it is to be understood that the invention encompasses all variations, combinations, and permutations in which one or more limitations, elements, clauses, descriptive terms, *etc.*, from one or more of the claims or from relevant portions of the description is introduced into another claim. For example, any claim that is dependent on another claim can be modified to include one or more limitations found in any other claim that is dependent on the same base claim. Furthermore, where the claims recite a composition, it is to be understood that methods of using the composition for any of the
purposes disclosed herein are included, and methods of making the composition according to any of the methods of making disclosed herein or other methods known in the art are included, unless otherwise indicated or unless it would be evident to one of ordinary skill in the art that a contradiction or inconsistency would arise.

[00179] Where elements are presented as lists, e.g., in Markush group format, it is to be understood that each subgroup of the elements is also disclosed, and any element(s) can be removed from the group. It is also noted that the term "comprising" is intended to be open and permits the inclusion of additional elements or steps. It should be understood that, in general, where the invention, or aspects of the invention, is/are referred to as comprising particular elements, features, steps, etc., certain embodiments of the invention or aspects of the invention consist, or consist essentially of, such elements, features, steps, etc. For purposes of simplicity those embodiments have not been specifically set forth in haec verba herein. Thus for each embodiment of the invention that comprises one or more elements, features, steps, etc., the invention also provides embodiments that consist or consist essentially of those elements, features, steps, etc.

[00180] Where ranges are given, endpoints are included. Furthermore, it is to be understood that unless otherwise indicated or otherwise evident from the context and/or the understanding of one of ordinary skill in the art, values that are expressed as ranges can assume any specific value within the stated ranges in different embodiments of the invention, to the tenth of the unit of the lower limit of the range, unless the context clearly dictates otherwise. It is also to be understood that unless otherwise indicated or otherwise evident from the context and/or the understanding of one of ordinary skill in the art, values expressed as ranges can assume any subrange within the given range, wherein the endpoints of the subrange are expressed to the same degree of accuracy as the tenth of the unit of the lower limit of the range.

[00181] In addition, it is to be understood that any particular embodiment of the present invention may be explicitly excluded from any one or more of the claims. Where ranges are given, any value within the range may explicitly be excluded from any one or more of the claims. Any embodiment, element, feature, application, or aspect of the compositions and/or methods of the invention, can be excluded from any one or more claims. For purposes of brevity, all of the embodiments in which one or more elements, features, purposes, or aspects is excluded are not set forth explicitly herein.

[00182] All publications, patents and sequence database entries mentioned herein, including those items listed above, are hereby incorporated by reference in their entirety as if
each individual publication or patent was specifically and individually indicated to be incorporated by reference. In case of conflict, the present application, including any definitions herein, will control.
CLAIMS

What is claimed is:

1. A Cas9 protein comprising an intein.

2. The Cas9 protein of claim 1, wherein an activity of the Cas9 protein is disrupted by the intein, and wherein the disrupted activity is restored upon excision of the intein from the Cas9 protein,

3. The Cas9 protein of claim 1 or claim 2, wherein the nuclease activity of the Cas9 protein is disrupted.

4. The Cas9 protein of claim 3, wherein the Cas9 protein exhibits no or minimal nuclease activity prior to excision of the intein.

5. The Cas9 protein of claim 4, wherein the nuclease activity of the Cas9 protein increases at least 2-fold, at least 3-fold, at least 4-fold, at least 5-fold, at least 6-fold, at least 7-fold, at least 8-fold, at least 9-fold, at least 10-fold, at least 15-fold, at least 20-fold, at least 30-fold, at least 40-fold, or at least 50-fold upon excision of the intein.

6. The Cas9 protein of any one of claims 1-5, wherein the Cas9 protein is capable of binding a guide RNA (gRNA) prior to excision of the intein.

7. The Cas9 protein of any one of claims 1-5, wherein the Cas9 protein has no or minimal gRNA binding activity prior to excision of the intein.

8. The Cas9 protein of any one of claims 1-7, wherein the Cas9 protein is capable of binding a gRNA and cleaving a target nucleic acid upon excision of the intein.

9. The Cas9 protein of claim 8, wherein the intein is inserted into the gRNA binding domain, the HNH nuclease domain, or the RuvC nuclease domain.
10. The Cas9 protein of any one of claims 1-9, wherein the Cas9 protein comprises a nuclease-inactivated Cas9 (dCas9) domain.

11. The Cas9 protein of any one of claims 1-10, wherein the Cas9 protein is a Cas9 nickase.

12. The Cas9 protein of any one of claims 1-11, wherein the intein replaces amino acid residue Cys80, Alal27, Thr146, Ser219, Thr333, Thr519, Cys574, Thr622, Ser701, Ala728, Thr995, Serl006, Serll54, Serll59, or Serl274 in the Cas9 polypeptide sequence set forth as SEQ ID NO:2, in the dCas9 polypeptide sequence set forth as SEQ ID NO:5, or in the Cas9 nickase polypeptide sequence set forth as SEQ ID NO:4, or an amino acid residue corresponding or homologous to any one of these residues; or wherein the intein replaces or is inserted at an amino acid residue that is within 5, within 10, within 15, or within 20 amino acid residues of Cys80, Alal27, Thr146, Ser219, Thr333, Thr519, Cys574, Thr622, Ser701, Ala728, Thr995, Serl006, Serll54, Serll59, or Serl274 in the Cas9 polypeptide sequence set forth as SEQ ID NO:2, in the dCas9 polypeptide sequence set forth as SEQ ID NO:5, or in the Cas9 nickase polypeptide sequence set forth as SEQ ID NO:4, or an amino acid residue corresponding or homologous to any one of these residues.

13. The Cas9 protein of any one of claims 1-11, wherein the intein replaces amino acid residue Alal27, Thr146, Ser219, Thr519, or Cys574 in the Cas9 polypeptide sequence set forth as SEQ ID NO:2, or in the dCas9 polypeptide sequence set forth as SEQ ID NO:5, or in the Cas9 nickase polypeptide sequence set forth as SEQ ID NO:4, or an amino acid residue corresponding or homologous to any one of these residues.

14. The Cas9 protein of any one of claims 1-13, wherein the intein is a ligand-dependent intein.

15. The Cas9 protein of any one of claims 1-14, wherein the intein comprises a ligand-binding domain.
16. The Cas9 protein of claim 15, wherein the ligand-binding domain comprises an estrogen-binding domain.

17. The Cas9 protein of claim 16, wherein the ligand binding domain is derived from the ligand binding domain of the estrogen receptor.

18. The Cas9 protein of claim 17, wherein the ligand-binding domain comprises an amino acid sequence as provided in residues 304-551 of the human estrogen receptor, or an estrogen-binding fragment or variant thereof.

19. The Cas9 protein of claim 18, wherein the estrogen-binding domain comprises the sequence amino acid sequence

```
NSLALS LTADQMVSA LLDACEPPIL YSEYDPTSPFSEA SMMGLL TNLADRELVHMINW
AKRVPGFVDLTLHDQAHLLEC AWLEILMIGLVRSMHEPGKLLFAPNLLLDRNQGK
CVEGMVEIFDMLLATTSSRFRMNLQGEFVCLKSII LNSGVYTFLSTLKSLEEKDH
IHRA LDKITDLIIHMAKAGLTLQQHQRLAQLLLILSHIRHMSNKGM ELYSMKYT
NVVPYDLLLL EMLDAHRLHA,
```

or wherein the estrogen-binding domain comprises the amino acid sequence

```
NSLALS LTADQMVSA LLDACEPPIL*YSEYD*PTSPFSEA SMMGLL TNLADRELVHMINW
AKRVPGFVDLTLHDQAHLLEC*AWLEILMIGLVRSMHEPGKLLFAPNLLLDRNQ
CVEGMVEIFDMLLATTSSRFRMNLQGEFVCLKSII LNSGVYTFLSTLKSLEEK
DHIRALDKITDLIIHMAKAGLTLQQHQRLAQLLLILSHIRHMSNKGM ELYSMKYT
*NVVPYDLLLL EMLDAHRLHA,
```

wherein at least one of the residues L*, D*, C*, or T* is mutated.

20. The Cas9 protein of claim 19, wherein the estrogen-binding domain comprises at least one of the following mutations: L*P, D*N, C*R, or T*K.

21. The Cas9 protein of any one of claims 1-20, wherein the intein comprises an amino acid sequence selected from the group consisting of SEQ ID NOs:7-14.

22. The Cas9 protein of any one of claims 14-21, wherein the ligand is selected from the group consisting of small molecules, peptides, proteins, polynucleotides, amino acids, and nucleotides.
23. The Cas9 protein of claim 22, wherein the ligand is a small molecule.

24. The Cas9 protein of any one of claims 14-23, wherein the ligand is 17β-estradiol, 17cc-ethynyl estradiol, tamoxifen, a tamoxifen analog, 3-hydroxytamoxifen (droloxifene), a tamoxifen metabolite, raloxifene, toremifene, ICT182, or ICT780.

25. The Cas9 protein of claim 24, wherein the ligand is 4-hydroxytamoxifen.

26. The Cas9 protein of any one of claims 10-25, wherein the Cas9 protein comprises: (i) a dCas9 domain; (ii) a ligand-dependent intein; and (iii) a recombinase catalytic domain.

27. The Cas9 protein of claim 26, wherein the recombinase catalytic domain is a monomer of the recombinase catalytic domain of Hin recombinase, Gin recombinase, or Tn3 recombinase.

28. The Cas9 protein of claim 26 or 27, wherein the Cas9 protein has no or minimal recombinase activity prior to excision of the intein.

29. The Cas9 protein of any one of claims 26-28, wherein the Cas9 protein is capable of binding a gRNA prior to excision of the intein.

30. The Cas9 protein of any one of claims 16-29, wherein the Cas9 protein has no or minimal RNA binding activity prior to excision of the intein.

31. The Cas9 protein of any one of claims 16-30, wherein the Cas9 protein is capable of binding a gRNA and hybridizing to a target nucleic acid upon excision of the intein.

32. The Cas9 protein of any one of claims 7-25, wherein the Cas9 protein is a Cas9 protein comprising three domains: (i) a dCas9 domain; (ii) a ligand-dependent intein; and (iii) a nucleic acid-editing domain.

33. The Cas9 protein of claim 32, wherein the nucleic acid-editing domain is a DNA-editing domain.
34. The Cas9 protein of claim 32 or 33, wherein the nucleic acid-editing domain is a
deaminase domain.

35. The Cas9 protein of any one of claims 32-34, wherein the deaminase is a cytidine
deaminase.

36. The Cas9 protein of claim 35, wherein the deaminase is an apolipoprotein B mRNA-
editing complex (APOBEC) family deaminase.

37. The Cas9 protein of claim 36, wherein the deaminase is an APOBEC 1 family
deaminase.

38. The Cas9 protein of claim 35, wherein the deaminase is an activation-induced
cytidine deaminase (AID).

39. The Cas9 protein of claim 35, wherein the deaminase is an ACF1/ASE deaminase.

40. The Cas9 protein of any one of claims 32-34, wherein the deaminase is an adenosine
deaminase.

41. The Cas9 protein of claim 40, wherein the deaminase is an ADAT family deaminase.

42. The Cas9 protein of any one of claims 32-41, wherein the Cas9 protein exhibits no or
minimal deaminase activity prior to excision of the intein.

43. The Cas9 protein of any one of claims 32-42, wherein the Cas9 protein is capable of
binding a gRNA prior to excision of the intein.

44. The Cas9 protein of any one of claims 32-42, wherein the Cas9 protein exhibits no or
minimal RNA binding activity prior to excision of the intein.

45. The Cas9 protein of any one of claims 32-44, wherein the Cas9 protein is capable of
binding a gRNA and hybridizing to a target nucleic acid upon excision of the intein.
46. The Cas9 protein of any one of claims 7-25, wherein the Cas9 protein comprises: (i) a dCas9 domain; (ii) a ligand-dependent intein; and (iii) a transcriptional activator domain.

47. The Cas9 protein of claim 46, wherein the transcriptional activator is selected from the group consisting of VP64, VP16, and p65.

48. The Cas9 protein of any one of claims 7-25, wherein the Cas9 protein comprises: (i) a dCas9 domain; (ii) a ligand-dependent intein; and (iii) a transcriptional repressor domain.

49. The Cas9 protein of claim 48, wherein the transcriptional repressor comprises a KRAB or SID domain.

50. The Cas9 protein of any one of claims 7-25, wherein the Cas9 protein comprises: (i) a dCas9 domain; (ii) a ligand-dependent intein; and (iii) an epigenetic modifier domain.

51. The Cas9 protein of claim 50, wherein the epigenetic modifier is selected from the group consisting of histone demethylase, histone methyltransferase, hydroxylase, histone deacetylase, and histone acetyltransferase.

52. The Cas9 fusion protein of claim 51, wherein the epigenetic modifier comprises the LSD1 histone demethylase or TET1 hydroxylase.

53. The Cas9 protein of any one of claims 1-52, wherein one or more domains of the Cas9 protein are separated by a peptide linker or a non-peptide linker.

54. The Cas9 protein of claim 53, wherein the peptide linker comprises an XTEN linker or an amino acid sequence comprising one or more repeats of the tri-peptide GGS.

55. A method for site-specific DNA cleavage comprising:
   (a) contacting a Cas9 protein of any one of claims 1-54 with a ligand, wherein binding of the ligand to the intein induces self-excision of the intein; and
   (b) contacting a DNA with the Cas9 protein, wherein the Cas9 protein is associated with a gRNA;
wherein self-excision of the intein from the Cas9 protein in step (a) allows the RNA-guided nuclease to cleave the DNA, thereby producing cleaved DNA.

56. The method of claim 55, wherein step (b) precedes step (a).

57. The method of claim 55 or 56, wherein the RNA-guided nuclease is able to bind the gRNA after excision of the intein.

58. A method for site-specific recombination between two DNA molecules, comprising:
   (a) contacting a first DNA with a first Cas9 protein of any one of claims 26-31, wherein the dCas9 domain of the first Cas9 protein binds a first gRNA that hybridizes to a region of the first DNA;
   (b) contacting the first DNA with a second Cas9 protein of any one of claims 26-31, wherein the dCas9 domain of the second Cas9 protein binds a second gRNA that hybridizes to a second region of the first DNA;
   (c) contacting a second DNA with a third Cas9 protein of any one of claims 26-31, wherein the dCas9 domain of the third Cas9 protein binds a third gRNA that hybridizes to a region of the second DNA; and
   (d) contacting the second DNA with a fourth Cas9 protein of any one of claims 26-31, wherein the dCas9 domain of the fourth Cas9 protein binds a fourth gRNA that hybridizes to a second region of the second DNA;

   wherein the binding of the Cas9 proteins in steps (a) - (d) results in the tetramerization of the recombinase catalytic domains of the Cas9 proteins, under conditions such that the DNAs are recombined.

59. The method of claim 58, wherein the gRNAs of steps (a) and (b) hybridize to opposing strands of the first DNA, and the gRNAs of steps (c) and (d) hybridize to opposing strands of the second DNA.

60. The method of claim 59, wherein the gRNAs of steps (a) and (b); and/or the gRNAs of steps (c) and (d) hybridize to regions of their respective DNAs that are no more than 10, no more than 15, no more than 20, no more than 25, no more than 30, no more than 40, no more than 50, no more than 60, no more than 70, no more than 80, no more than 90, or no more than 100 base pairs apart.
61. A method for site-specific recombination between two regions of a single DNA molecule, comprising:

(a) contacting the DNA with a first Cas9 protein of any one of claims 26-31, wherein the dCas9 domain if the first Cas9 protein binds a first gRNA that hybridizes to a region of the DNA;

(b) contacting the DNA with a second Cas9 protein of any one of claims 26-31, wherein the dCas9 domain of the second Cas9 protein binds a second gRNA that hybridizes to a second region of the DNA;

(c) contacting the DNA with a third Cas9 protein of any one of claims 26-31, wherein the dCas9 domain of the third Cas9 protein binds a third gRNA that hybridizes to a third region of the DNA;

(d) contacting the DNA with a fourth Cas9 protein of any one of claims 26-31, wherein the dCas9 domain of the fourth Cas9 protein binds a fourth gRNA that hybridizes to a fourth region of the DNA;

wherein the binding of the Cas9 proteins in steps (a) - (d) results in the tetramerization of the recombinase catalytic domains of the Cas9 proteins, under conditions such that the DNA is recombined.

62. The method of claim 61, wherein two of the gRNAs of steps (a) - (d) hybridize to the same strand of the DNA, and the other two gRNAs of steps (a)- (d) hybridize to the opposing strand of the DNA.

63. The method of claim 61 or 62, wherein the gRNAs of steps (a) and (b) hybridize to regions of the DNA that are no more than 50, no more than 60, no more than 70, no more than 80, no more than 90, or no more than 100 base pairs apart, and the gRNAs of steps (c) and (d) hybridize to regions of the DNA that are no more than 10, no more than 15, no more than 20, no more than 25, no more than 30, no more than 40, no more than 50, no more than 60, no more than 70, no more than 80, no more than 90, or no more than 100 base pairs apart.

64. The method of any one of claims 58-63, wherein the method further comprises contacting the Cas9 proteins with a ligand that induces self-excision of the intein.
65. The method of claim 64, wherein the Cas9 proteins are contacted with the ligand prior to forming a complex with a gRNA.

66. The method of claim 64, wherein the Cas9 proteins are contacted with the ligand after forming a complex with a gRNA.

67. A method of DNA editing, the method comprising contacting a DNA molecule with (a) a Cas9 protein of any one of claims 32-45; and (b) a gRNA targeting the Cas9 protein of (a) to a target nucleotide sequence of the DNA strand;

wherein the DNA molecule is contacted with the Cas9 protein and the gRNA in an amount effective and under conditions suitable for the deamination of a nucleotide base.

68. The method of claim 67, wherein the method further comprises contacting the Cas9 protein with a ligand that induces self-excision of the intein.

69. The method of claim 68, wherein the Cas9 protein is contacted with the ligand prior to forming a complex with a gRNA.

70. The method of claim 68, wherein the Cas9 protein is contacted with the ligand after forming a complex with a gRNA.

71. The method of claim 67, wherein the target DNA sequence comprises a sequence associated with a disease or disorder, and wherein the deamination of the nucleotide base results in a sequence that is not associated with a disease or disorder.

72. The method of claim 71, wherein the DNA sequence comprises a T→C or A→G point mutation associated with a disease or disorder, and wherein the deamination of the mutant C or G base results in a sequence that is not associated with a disease or disorder.

73. The method of claim 71 or 72, wherein the deamination corrects a point mutation in the sequence associated with the disease or disorder.
74. The method of claim 1 or 72, wherein the sequence associated with the disease or disorder encodes a protein, and wherein the deamination introduces a stop codon into the sequence associated with the disease or disorder, resulting in a truncation of the encoded protein.

75. The method of any one of claims 67-72, wherein the deamination corrects a point mutation in the PIK3CA gene, thus correcting an H1047R and/or a A3140G mutation.

76. The method of any one of claims 71-75, wherein the contacting is in vivo in a subject having or diagnosed with the disease or disorder.

77. The method of any one of claims 71-76, wherein the disease or disorder is a cancer.

78. A method for transcriptional activation, comprising contacting a DNA molecule comprising a gene with
   (a) a Cas9 protein of claims 46 or 47; and
   (b) a gRNA targeting the Cas9 protein of (a) to a target nucleotide sequence of the DNA strand;

   wherein the DNA molecule is contacted with the Cas9 protein and the gRNA in an amount effective and under conditions suitable for the transcriptional activation of the gene.

79. A method for transcriptional repression, comprising contacting a DNA molecule comprising a gene with
   (a) a Cas9 protein of claims 48 or 49; and
   (b) a gRNA targeting the Cas9 protein of (a) to a target nucleotide sequence of the DNA strand;

   wherein the DNA molecule is contacted with the Cas9 protein and the gRNA in an amount effective and under conditions suitable for the transcriptional repression of the gene.

80. A method for epigenetic modification, comprising contacting a DNA molecule with
   (a) a Cas9 protein of any one of claims 50-52; and
   (b) a gRNA targeting the Cas9 protein of (a) to a target nucleotide sequence of the DNA strand;
wherein the DNA molecule is contacted with the Cas9 protein and the gRNA in an amount effective and under conditions suitable for the epigenetic modification of the DNA.

81. The method of any one of claims 78-80, wherein the method further comprises contacting the Cas9 protein with a ligand that induces self-excision of the intein.

82. The method of any one of claims 78-80, wherein the Cas9 protein is contacted with the ligand prior to forming a complex with a gRNA.

83. The method of any one of claims 78-80, wherein the Cas9 protein is contacted with the ligand after forming a complex with a gRNA.

84. The method of any one of claims 55-83, wherein the DNA is in a cell.

85. The method of claim 84, wherein the cell is a eukaryotic cell.

86. The method of claim 85, wherein the cell is a human cell.

87. The method of claim 85 or 86, wherein the cell is in a subject.

88. The method of claim 87, wherein the subject is a human.

89. A polynucleotide encoding a Cas9 protein of any one of claims 1-54.

90. A vector comprising a polynucleotide of claim 89.

91. A cell comprising a genetic construct for expressing a Cas9 protein of any one of claims 1-54.

92. A kit comprising a Cas9 protein of any one of claims 1-54.

93. A kit comprising a polynucleotide of claim 89 and/or a vector of claim 90.
94. A kit comprising a vector for recombinant protein expression, wherein the vector comprises a polynucleotide encoding a Cas9 protein of any one of claims 1-54.

95. A kit comprising a cell that comprises a genetic construct for expressing a Cas9 protein of any one of claims 1-54.
1: - 4-HT, Wild-type Cas9 (500 ng)
2: - 4-HT, Wild-type Cas9 (80 ng)
3: - 4-HT, Cas9:Intein (37R3-2 replacing S219; 500 ng)
4: - 4-HT, Cas9:Intein (37R3-2 replacing C574; 500 ng)
5: + 4-HT, Wild-type Cas9 (500 ng)
6: + 4-HT, Wild-type Cas9 (80 ng)
7: + 4-HT, Cas9:Intein (37R3-2 replacing S219; 500 ng)
8: + 4-HT, Cas9:Intein (37R3-2 replacing C574; 500 ng)

Figure 2
Figure 3B

Figure 3C
Figure 4A

Figure 4B
Figure 4C

Figure 4D
Figure 5
Figure 6A

Figure 6B
Figure 7

- + 4-HT intein-Cas9(S219)
- + 4-HT intein-Cas9(C574)
- + 4-HT wt Cas9

Indel modification frequency (%)
Figure 8A

Figure 8B

Figure 8C
Figure 9

- + 4-HT intein-Cas9 (S219)
- + 4-HT intein-Cas9 (C574)
- + 4-HT wt Cas9 (500 ng)
- + 4-HT wt Cas9 (260 ng)
- + 4-HT wt Cas9 (140 ng)
- + 4-HT wt Cas9 (80 ng)
- + 4-HT wt Cas9 (50 ng)
Figure 10A

- + 4-HT intein-Cas9(S219)
- + 4-HT intein-Cas9(C574)
- + 4-HT wt Cas9 (500 ng)
- + 4-HT wt Cas9 (260 ng)
- + 4-HT wt Cas9 (140 ng)
- + 4-HT wt Cas9 (80 ng)
- + 4-HT wt Cas9 (50 ng)

Figure 10B

- + 4-HT intein-Cas9(S219)
- + 4-HT intein-Cas9(C574)
- + 4-HT wt Cas9 (500 ng)
- + 4-HT wt Cas9 (260 ng)
- + 4-HT wt Cas9 (140 ng)
- + 4-HT wt Cas9 (80 ng)
- + 4-HT wt Cas9 (50 ng)
Figure 12A

Figure 12B