(12) STANDARD PATENT (11) Application No. **AU 2014362208 B2**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Cas variants for gene editing

(51) International Patent Classification(s)
C12N 9/78 (2006.01)

(21) Application No: **2014362208** (22) Date of Filing: **2014.12.12**

(87) WIPO No: **WO15/089406**

(30) Priority Data

(31) Number	(32) Date	(33) Country
14/326,318	**2014.07.08**	**US**
61/980,333	**2014.04.16**	**US**
14/326,303	**2014.07.08**	**US**
14/326,140	**2014.07.08**	**US**
14/326,290	**2014.07.08**	**US**
61/915,386	**2013.12.12**	**US**
14/326,109	**2014.07.08**	**US**
14/326,269	**2014.07.08**	**US**
14/325,815	**2014.07.08**	**US**

(43) Publication Date: **2015.06.18**
(44) Accepted Journal Date: **2021.02.11**

(71) Applicant(s)
President and Fellows of Harvard College

(72) Inventor(s)
Liu, David R.;Komor, Alexis Christine

(74) Agent / Attorney
Pizzeys Patent and Trade Mark Attorneys Pty Ltd, PO Box 291, WODEN, ACT, 2606, AU

(56) Related Art
WO 2014186686 A2
PRASHANT MALI ET AL, "Cas9 as a versatile tool for engineering biology", NATURE METHODS, NATURE PUBLISHING GROUP, GB, vol. 10, no. 10, doi:10.1038/NMETH.2649, ISSN 1548-7105, (2013-10-01), pages 957 - 963, (2013-09-27)
WO 2014/089290 A1
LEI S. QI ET AL, "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression", CELL, (2013-02-01), vol. 152, no. 5, doi:10.1016/j.cell.2013.02.022, ISSN 0092-8674, pages 1173 - 1183

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 June 2015 (18.06.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/089406 A1

(51) International Patent Classification:
C12N 9/78 (2006.01)

(21) International Application Number:
PCT/US2014/070038

(22) International Filing Date:
12 December 2014 (12.12.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

61/915,386	12 December 2013 (12.12.2013)	US
61/980,333	16 April 2014 (16.04.2014)	US
14/326,318	8 July 2014 (08.07.2014)	US
14/326,140	8 July 2014 (08.07.2014)	US
14/326,303	8 July 2014 (08.07.2014)	US
14/326,290	8 July 2014 (08.07.2014)	US
14/326,269	8 July 2014 (08.07.2014)	US
14/326,109	8 July 2014 (08.07.2014)	US
14/325,815	8 July 2014 (08.07.2014)	US

(71) Applicant: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US).

(72) Inventors: **LIU, David, R.**; 3 Whitman Circle, Lexington, MA 02420 (US). **KOMOR, Alexis, Christine**; 1000 Cordova St., #109, Pasadena, CA 91106 (US).

(74) Agent: **BAKER, C., Hunter**; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210-2206 (US).

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

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

Published:

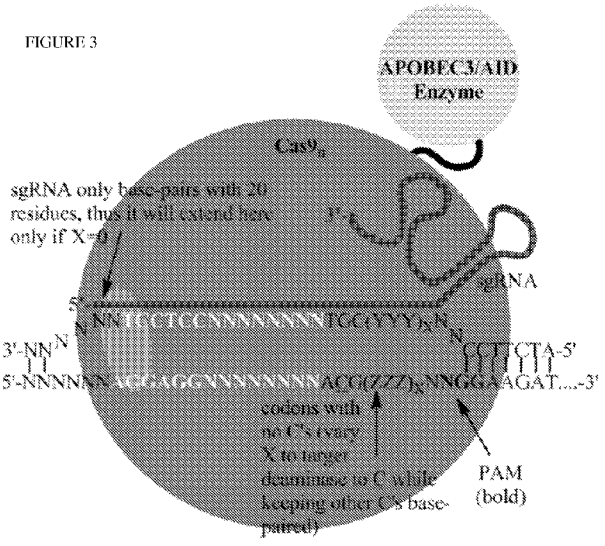
— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

WO 2015/089406 A1

(54) Title: CAS VARIANTS FOR GENE EDITING

FIGURE 3



(57) Abstract: Some aspects of this disclosure provide strategies, systems, reagents, methods, and kits that are useful for the targeted editing of nucleic acids, including editing a single site within the genome of a cell or subject, *e.g.*, within the human genome. In some embodiments, fusion proteins of Cas9 and nucleic acid editing enzymes or enzyme domains, *e.g.* , deaminase domains, are provided. In some embodiments, methods for targeted nucleic acid editing are provided. In some embodiments, reagents and kits for the generation of targeted nucleic acid editing proteins, *e.g.*, fusion proteins of Cas9 and nucleic acid editing enzymes or domains, are provided.

CAS VARIANTS FOR GENE EDITING

RELATED APPLICATION

[0001] This application claims priority under 35 U.S.C. § 119(e) to U.S. provisional patent application, U.S.S.N. 61/915,386 filed December 12, 2013, and U.S. provisional patent application, U.S.S.N. 61/980,333 filed April 16, 2014; and also claims priority under 35 U.S.C. § 120 to U.S. patent applications, U.S.S.N. 14/325,815, 14/326,109, 14/326,140, 14/326,269, 14/326,290, 14/326,318, and 14/326,303, all filed on July 8, 2014; each of which is incorporated herein by reference.

GOVERNMENT SUPPORT

[0002] This invention was made with U.S. Government support under grant HR0011-11-2-0003 awarded by the Defense Advanced Research Projects Agency (DARPA), grant GM095501 awarded by the National Institutes of Health (NIH), and grant N66001-12-C-4207 awarded by the Space and Naval Warfare Systems Center (SPAWAR). The Government has certain rights in this invention.

BACKGROUND OF THE INVENTION

[0003] Targeted editing of nucleic acid sequences, for example, the introduction of a specific modification into genomic DNA, is a highly promising approach for the study of gene function and also has the potential to provide new therapies for human genetic diseases.[1] An ideal nucleic acid editing technology possesses three characteristics: (1) high efficiency of installing the desired modification; (2) minimal off-target activity; and (3) the ability to be programmed to edit precisely any site in a given nucleic acid, *e.g.*, any site within the human genome.[2] Current genome engineering tools, including engineered zinc finger nucleases (ZFNs),[3] transcription activator like effector nucleases (TALENs),[4] and most recently, the RNA-guided DNA endonuclease Cas9,[5] effect sequence-specific DNA cleavage in a genome. This programmable cleavage can result in mutation of the DNA at the cleavage site via non-homologous end joining (NHEJ) or replacement of the DNA surrounding the cleavage site via homology-directed repair (HDR).[6,7]

[0004] One drawback to the current technologies is that both NHEJ and HDR are stochastic processes that typically result in modest gene editing efficiencies as well as unwanted gene alterations that can compete with the desired alteration.[8] Since many genetic diseases in principle can be treated by effecting a specific nucleotide change at a specific

location in the genome (for example, a C to T change in a specific codon of a gene associated with a disease),[9] the development of a programmable way to achieve such precision gene editing would represent both a powerful new research tool, as well as a potential new approach to gene editing-based human therapeutics.

SUMMARY OF THE INVENTION

[0005] The clustered regularly interspaced short palindromic repeat (CRISPR) system is a recently discovered prokaryotic adaptive immune system[10] that has been modified to enable robust and general genome engineering in a variety of organisms and cell lines.[11] CRISPR-Cas (CRISPR associated) systems are protein-RNA complexes that use an RNA molecule (sgRNA) as a guide to localize the complex to a target DNA sequence *via* base-pairing.[12] In the natural systems, a Cas protein then acts as an endonuclease to cleave the targeted DNA sequence.[13] The target DNA sequence must be both complementary to the sgRNA, and also contain a "protospacer-adjacent motif" (PAM) dinucleotide at the 3′-end of the complementary region in order for the system to function (Figure 1).[14] Among the known Cas proteins, *S. pyogenes* Cas9 has been mostly widely used as a tool for genome engineering.[15] This Cas9 protein is a large, multi-domain protein containing two distinct nuclease domains. Point mutations can be introduced into Cas9 to abolish nuclease activity, resulting in a dead Cas9 (dCas9) that still retains its ability to bind DNA in a sgRNA-programmed manner.[16] In principle, when fused to another protein or domain, dCas9 can target that protein to virtually any DNA sequence simply by co-expression with an appropriate sgRNA.

[0006] The potential of the dCas9 complex for genome engineering purposes is immense. Its unique ability to bring proteins to specific sites in a genome programmed by the sgRNA in theory can be developed into a variety of site-specific genome engineering tools beyond nucleases, including transcriptional activators, transcriptional repressors, histone-modifying proteins, integrases, and recombinases.[11] Some of these potential applications have recently been implemented through dCas9 fusions with transcriptional activators to afford RNA-guided transcriptional activators,[17,18] transcriptional repressors,[16,19,20] and chromatin modification enzymes.[21] Simple co-expression of these fusions with a variety of sgRNAs results in specific expression of the target genes. These seminal studies have paved the way for the design and construction of readily programmable sequence-specific effectors for the precise manipulation of genomes.

[0007] Significantly, 80-90% of protein mutations responsible for human disease arise from the substitution, deletion, or insertion of only a single nucleotide.[6] No genome engineering tools, however, have yet been developed that enable the manipulation of a single nucleotide in a general and direct manner. Current strategies for single-base gene correction include engineered nucleases (which rely on the creation of double-strand breaks, DSBs ,followed by stochastic, inefficient homology-directed repair, HDR), and DNA-RNA chimeric oligonucleotides.[22] The latter strategy involves the design of a RNA/DNA sequence to base pair with a specific sequence in genomic DNA except at the nucleotide to be edited. The resulting mismatch is recognized by the cell's endogenous repair system and fixed, leading to a change in the sequence of either the chimera or the genome. Both of these strategies suffer from low gene editing efficiencies and unwanted gene alterations, as they are subject to both the stochasticity of HDR and the competition between HDR and non-homologous end-joining, NHEJ.[23-25] HDR efficiencies vary according to the location of the target gene within the genome,[26] the state of the cell cycle,[27] and the type of cell/tissue.[28] The development of a direct, programmable way to install a specific type of base modification at a precise location in genomic DNA with enzyme-like efficiency and no stochasticity would therefore represent a powerful new approach to gene editing-based research tools and human therapeutics.

[0008] Some aspects of this disclosure provide strategies, systems, reagents, methods, and kits that are useful for the targeted editing of nucleic acids, including editing a single site within a subject's genome, *e.g.*, the human genome. In some embodiments, fusion proteins of Cas9 and nucleic acid editing enzymes or enzyme domains, *e.g.*, deaminase domains, are provided. In some embodiments, methods for targeted nucleic acid editing are provided. In some embodiments, reagents and kits for the generation of targeted nucleic acid editing proteins, *e.g.*, fusion proteins of Cas9 and nucleic acid editing enzymes or domains, are provided.

[0009] Some aspects of this disclosure provide fusion proteins comprising (i) a nuclease-inactive CAS9 domain; and (ii) a nucleic acid-editing domain. In some embodiments, the nucleic acid-editing domain is a DNA-editing domain. In some embodiments, the nucleic-acid-editing domain is a deaminase domain. In some embodiments, the deaminase is a cytidine deaminase. In some embodiments, the deaminase is an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase. In some embodiments, the deaminase is an APOBEC1 family deaminase. In some embodiments, the deaminase is an activation-induced cytidine deaminase (AID). In some embodiments, the

deaminase is an ACF1/ASE deaminase. In some embodiments, the deaminase is an adenosine deaminase. In some embodiments, the deaminase is an ADAT family deaminase. In some embodiments, the nucleic-acid-editing domain is fused to the N-terminus of the CAS9 domain. In some embodiments, the nucleic-acid-editing domain is fused to the C-terminus of the CAS9 domain. In some embodiments, the CAS9 domain and the nucleic-acid-editing domain are fused via a linker. In some embodiments, the linker comprises a $(GGGGS)_n$ (SEQ ID NO: 91), a $(G)_n$, an $(EAAAK)_n$ (SEQ ID NO: 5), a $(GGS)_n$, an SGSETPGTSESATPES (SEQ ID NO: 93) motif (see, *e.g.*, Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* **2014;** 32(6): 577-82; the entire contents are incorporated herein by reference), or an $(XP)_n$ motif, or a combination of any of these, wherein n is independently an integer between 1 and 30.

[0010] Some aspects of this disclosure provide methods for DNA editing. In some embodiments, the methods comprise contacting a DNA molecule with (a) a fusion protein comprising a nuclease-inactive Cas9 domain and a deaminase domain; and (b) an sgRNA 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 sgRNA in an amount effective and under conditions suitable for the deamination of a nucleotide base. 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 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. 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, 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.

[0011] Some aspects of this disclosure provide a reporter construct for detecting nucleic-acid-editing activity of a Cas9:DNA-editing domain fusion protein. In some embodiments, the construct comprises (a) a reporter gene comprising a target site for the Cas9 DNA-editing protein, wherein targeted DNA editing results in an increase in expression of the reporter gene; and (b) a promoter sequence that controls expression of the reporter gene. In some embodiments, the construct further comprises (c) a sequence encoding an sgRNA targeting the Cas9 DNA-editing protein to the target site of the reporter gene, wherein expression of the sgRNA is independent of the expression of the reporter gene. In some embodiments, the target site of the reporter gene comprises a premature stop codon, and wherein targeted DNA editing of the template strand by the Cas9 DNA-editing protein results in a conversion of the premature stop codon to a codon encoding an amino acid residue. In some embodiments, the reporter gene encodes a luciferase, a fluorescent protein, or an antibiotic resistance marker.

[0012] Some aspects of this disclosure provide kits comprising a nucleic acid construct that comprises a sequence encoding a nuclease-inactive Cas9 sequence, a sequence comprising a cloning site positioned to allow cloning of a sequence encoding a nucleic acid-editing enzyme or enzyme domain in-frame with the Cas9-encoding sequence, and, optionally, a sequence encoding a linker positioned between the Cas9 encoding sequence and the cloning site. In addition, in some embodiments, the kit comprises suitable reagents, buffers, and/or instructions for in-frame cloning of a sequence encoding a nucleic acid-editing enzyme or enzyme domain into the nucleic acid construct to generate a Cas9 nucleic acid editing fusion protein. In some embodiments, the sequence comprising the cloning site is N-terminal of the Cas9 sequence. In some embodiments, the sequence comprising the cloning site is C-terminal of the Cas9 sequence. In some embodiments, the encoded linker comprises a $(GGGGS)_n$ (SEQ ID NO: 91), a $(G)_n$, an $(EAAAK)_n$(SEQ ID NO: 5), a $(GGS)_n$, an SGSETPGTSESATPES (SEQ ID NO: 93) motif (see, *e.g.*, Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* **2014;** 32(6): 577-82; the entire contents are incorporated herein by reference), or an $(XP)_n$ motif, or a combination of any of these, wherein n is independently an integer between 1 and 30.

[0013] Some aspects of this disclosure provide kits comprising a fusion protein comprising a nuclease-inactive Cas9 domain and a nucleic acid-editing enzyme or enzyme domain, and, optionally, a linker positioned between the Cas9 domain and the nucleic acid-editing enzyme or enzyme domain. In addition, in some embodiments, the kit comprises

suitable reagents, buffers, and/or instructions for using the fusion protein, *e.g.*, for *in vitro* or *in vivo* DNA or RNA editing. In some embodiments, the kit comprises instructions regarding the design and use of suitable sgRNAs for targeted editing of a nucleic acid sequence.

[0014] The summary above is meant to illustrate, in a non-limiting manner, some of the embodiments, advantages, features, and uses of the technology disclosed herein. Other embodiments, advantages, features, and uses of the technology disclosed herein will be apparent from the Detailed Description, the Drawings, the Examples, and the Claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] **Figure 1.** The Cas9/sgRNA-DNA complex. The 3′ end of the sgRNA forms a ribonucleoprotein complex with the Cas9 nuclease, while the 20 nt 5′ end of the sgRNA recognizes its complementary stretch of DNA. DNA binding requires the 3-nt PAM sequence 5′ to the target DNA. In the case of wtCas9, double-strand DNA cleavage occurs 3 nt from the PAM to produce blunt ends (shown by the arrows). It should be noted that the size of the bubble is unknown.

[0016] **Figure 2.** Crystal structure of the catalytic domain of APOBEC3G (PDB ID 3E1U). The core secondary structure, which is believed to be conserved among the entire family, consists of a five-stranded β-sheet (arrows) flanked by six α-helices. The active center loop (active site loop), is believed to be responsible for determining deamination specificity. The Zn^{2+} responsible for catalytic activity is shown as a sphere. Sequences correspond, from top to bottom, to SEQ ID NOs: 97-98.

[0017] **Figure 3.** Design of luciferase-based reporter assay. The sgRNA will be varied to target numerous sequences that correspond to regions prior to and including the luciferase gene in order to target the mutated start codon (C residue underlined). A "buffer" region will be added between the start codon and the luciferase gene to include codons of only A's and T's (shown as $(ZZZ)_X$). The Shine-Dalgarno sequence is indicated. In some embodiments, it is preferable to keep all C's base-paired to prevent off-target effects.

[0018] **Figure 4.** Deaminase assay. Sequences correspond, from top to bottom, to SEQ ID NOs: 99-105.

[0019] **Figure 5.** SDS PAGE gel of ssDNA edited by Cas9-APOBEC1 fusion proteins.

DEFINITIONS

[0020] 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.

[0021] The term "Cas9" or "Cas9 nuclease" refers to an RNA-guided nuclease comprising a Cas9 protein, or a fragment thereof (*e.g.*, a protein comprising an active or inactive DNA cleavage domain of Cas9, and/or the gRNA binding domain of Cas9). A Cas9 nuclease is also referred to sometimes as a casn1 nuclease or a CRISPR (clustered regularly interspaced short palindromic repeat)-associated nuclease. CRISPR is an adaptive immune system that provides protection against mobile genetic elements (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 RNAs. However, single guide RNAs ("sgRNA", or simply "gNRA") can be engineered so as to incorporate aspects of both the crRNA and tracrRNA into a single RNA species. 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 *et al.*, 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., White J., Yuan X., Clifton S.W., Roe B.A., 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 C.M., 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, a Cas9 nuclease has an inactive (*e.g.*, an inactivated) DNA cleavage domain.

[0022] A nuclease-inactivated Cas9 protein may interchangeably be referred to as a "dCas9" protein (for nuclease-"dead" Cas9). Methods for generating a Cas9 protein (or a fragment thereof) having an inactive DNA cleavage domain are known (See, *e.g.*, Jinek *et al., Science.* 337:816-821(2012); Qi *et al.,* "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression" (2013) *Cell.* 28;152(5):1173-83, the entire contents of each of which are incorporated herein by reference). For example, the DNA cleavage domain of Cas9 is known to include two subdomains, the HNH nuclease subdomain and the RuvC1 subdomain. The HNH subdomain cleaves the strand complementary to the gRNA, whereas the RuvC1 subdomain cleaves the non-complementary strand. Mutations within these subdomains can silence the nuclease activity of Cas9. For example, the mutations D10A and H841A completely inactivate the nuclease activity of *S. pyogenes* Cas9 (Jinek *et al., Science.* 337:816-821(2012); Qi *et al., Cell.* 28;152(5):1173-83 (2013). In some embodiments, proteins comprising fragments of Cas9 are provided. For example, in some embodiments, a protein comprises one of two Cas9 domains: (1) the gRNA binding domain of Cas9; or (2) the DNA cleavage domain of Cas9. In some embodiments, proteins comprising Cas9 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 is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% 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), 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 96% identical, at least about 97% 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 Sequence: NC_017053.1, SEQ ID NO:1 (nucleotide); SEQ ID NO:2 (amino acid)).

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGATGATTAT
AAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAAATCTTATAGGGGCT
CTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGG
AAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGA
CTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAA
GTTGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGCAGATTCTACTGATAAAGCGGAT
TTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAAT
CCTGATAATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAATCTACAATCAATTATTTGAAGAAAACCCT
ATTAACGCAAGTAGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTC
ATTGCTCAGCTCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGAATCTCATTGCTTTGTCATTGGGATTGACCCCT
AATTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTA
GATAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATT
TTACTTTCAGATATCCTAAGAGTAAATAGTGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAGCGCTAC
GATGAACATCATCAAGACTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATC
TTTTTTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTT
ATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGC
AAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGA
CAAGAAGACTTTTATCCATTTTTAAAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTAT
TATGTTGGTCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCA
TGGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGATAAA
AATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACA
AAGGTCAAATATGTTACTGAGGGAATGCGAAAACCAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGAT
TTACTCTTCAAAACAAATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTT
GATAGTGTTGAAATTTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGCGCCTACCATGATTTGCTAAAAATT
ATTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTA
TTTGAAGATAGGGGGATGATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAACAG
CTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCT
GGCAAAACAATATTAGATTTTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGAT
AGTTTGACATTTAAAGAAGATATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGATTGCT
AACTTAGCTGGCAGTCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAATTGTTGATGAACTGGTCAAAGTA
ATGGGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAAT
TCGCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTT
GAAAATACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTACAAAATGGAAGAGACATGTATGTGGACCAA
GAATTAGATATTAATCGTTTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCATTAAAGACGATTCA
ATAGACAATAAGGTACTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTC
AAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACG
AAAGCTGAACGTGGAGGTTTGAGTGAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAA
ATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCGA
GAGGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGT
GAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATAT
CCAAAACTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAGTCTGAG
CAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACA
CTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAAA
GGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAG
ACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGG
GATCCAAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAA
GGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAA
AATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAATAT
AGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTG
GCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGAT
AACGAACAAAAACAATTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTT
TCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCA
ATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTTTAAATAT
TTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCC
ATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAGGAGGTGACTGA (SEQ ID NO:1)

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEATRLKRTARRRYTRR
KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLADSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENL
IAQLPGEKRNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKF
IKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKV
MGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT
KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVR
EINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDW
DPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKY
SLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF
SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQS
ITGLYETRIDLSQLGGD (SEQ ID NO:2)

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

[0023] In some embodiments, wild type Cas9 corresponds to, or comprises SEQ ID NO:3 (nucleotide) and/or SEQ ID NO: 4 (amino acid):

ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCCGTTGGATGGGCTGTCATAACCGATGAATAC
AAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTCGATTAAAAAGAATCTTATCGGTGCC
CTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGC
AAGAACCGAATATGTTACTTACAAGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGT
TTGGAAGAGTCCTTCCTTGTCGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGATGAG
GTGGCATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGGAC
CTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAAT
CCGGACAACTCGGATGTCGACAAACTGTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCT
ATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTG
ATCGCACAATTACCCGGAGAGAAGAAAAATGGGTTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCA
AATTTTAAGTCGAACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTC
GACAATCTACTGGCACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATC
CTCCTATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTAC
GATGAACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATA
TTCTTTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTT
ATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAAAACTCAATCGCGAAGATCTACTGCGA
AAGCAGCGGACTTTCGACAACGGTAGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGG
CAGGAGGATTTTTATCCGTTCCTCAAAGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTAC
TATGTGGGACCCCTGGCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCA
TGGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACCAACTTTGACAAG
AATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTTCACAGTGTACAATGAACTCACG
AAAGTTAAGTATGTCACTGAGGGCATGCGTAAACCCGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGAT
CTGTTATTCAAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTC
GATTCTGTCGAGATCTCCGGGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAAGATA
ATTAAAGATAAGGACTTCCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTGTTGACTCTTACCCTC
TTTGAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCTCACCTGTTCGACGATAAGGTTATGAAACAG
TTAAAGAGGCGTCGCTATACGGGCTGGGGACGATTGTCGCGGAAACTTATCAACGGGATAAGAGACAAGCAAAGT
GGTAAAACTATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGAC
TCTTTAACCTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATATTGCG
AATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTC
ATGGGACGTCACAAACCGGAAAACATTGTAATCGAGATGGCACGCGAAAATCAAACGACTCAGAAGGGGCAAAAA
AACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGGTATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCT
GTGGAAAATACCCAATTGCAGAACGAGAAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGAT
CAGGAACTGGACATAAACCGTTTATCTGATTACGACGTCGATCACATTGTACCCCAATCCTTTTTGAAGGACGAT
TCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCGAGGGAAAAGTGACAATGTTCCAAGCGAGGAAGTC
GTAAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAAATGCGAAACTGATAACGCAAAGAAAGTTCGATAACTTA

ACTAAAGCTGAGAGGGGTGGCTTGTCTGAACTTGACAAGGCCGGATTTATTAAACGTCAGCTCGTGGAAACCCGC
CAAATCACAAAGCATGTTGCACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATT
CGGGAAGTCAAAGTAATCACTTTAAAGTCAAAATTGGTGTCGGACTTCAGAAAGGATTTTCAATTCTATAAAGTT
AGGGAGATAAATAACTACCACCATGCGCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTAAGAAA
TACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCCGTAAGATGATCGCGAAAAGC
GAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAATC
ACTCTGGCAAACGGAGAGATACGCAAACGACCTTTAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGAT
AAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTG
CAGACCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGAC
TGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCTATTCTGTCCTAGTAGTGGCAAAAGTTGAG
AAGGGAAAATCCAAGAAACTGAAGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAA
AAGAACCCCATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCATAATTAAACTACCAAAG
TATAGTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAGCTTCAAAAGGGGAACGAA
CTCGCACTACCGTCTAAATACGTGAATTTCCTGTATTTAGCGTCCCATTACGAGAAGTTGAAAGGTTCACCTGAA
GATAACGAACAGAAGCAACTTTTTGTTGAGCAGCACAAACATTATCTCGACGAAATCATAGAGCAAATTTCGGAA
TTCAGTAAGAGAGTCATCCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGGGATAAA
CCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGCATTCAAG
TATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGGAGGTGCTAGACGCGACACTGATTCACCAA
TCCATCACGGGATTATATGAAACTCGGATAGATTTGTCACAGCTTGGGGGTGACGGATCCCCCAAGAAGAAGAGG
AAAGTCTCGAGCGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGAC
AAGGCTGCAGGA (SEQ ID NO:3)

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR
KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENL
IAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKF
IKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVD
QELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD
WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISE
FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ
SITGLYETRIDLSQLGGD (SEQ ID NO:4)
(single underline: HNH domain; double underline: RuvC domain)

[0024] In some embodiments, dCas9 corresponds to, or comprises in part or in whole, a Cas9 amino acid sequence having one or more mutations that inactivate the Cas9 nuclease activity. For example, in some embodiments, a dCas9 domain comprises D10A and/or H820A mutation.

dCas9 (D10A and H840A):

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR
KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENL
IAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKF
IKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVD
QELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD
WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISE
FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ
SITGLYETRIDLSQLGGD (SEQ ID NO: 34)
(single underline: HNH domain; double underline: RuvC domain)

[0025] In other embodiments, dCas9 variants having mutations other than D10A and H820A are provided, which *e.g.*, result in nuclease inactivated Cas9 (dCas9). Such mutations, by way of example, include other amino acid substitutions at D10 and H820, or other substitutions within the nuclease domains of Cas9 (*e.g.*, substitutions in the HNH nuclease subdomain and/or the RuvC1 subdomain). In some embodiments, variants or homologues of dCas9 (*e.g.*, variants of SEQ ID NO: 34) 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: 34. In some embodiments, variants of dCas9 (*e.g.*, variants of SEQ ID NO: 34) are provided having amino acid sequences which are shorter, or longer than SEQ ID NO: 34, 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.

[0026] In some embodiments, Cas9 fusion proteins as provided herein comprise the full-length amino acid of a Cas9 protein, *e.g.*, one of the sequences provided above. In other embodiments, however, fusion proteins as provided herein do not comprise a full-length Cas9 sequence, but only a fragment thereof. For example, in some embodiments, a Cas9 fusion protein provided herein comprises a Cas9 fragment, wherein the fragment binds crRNA and tracrRNA or sgRNA, but does not comprise a functional nuclease domain, *e.g.*, in that it comprises only a truncated version of a nuclease domain or no nuclease domain at all. Exemplary amino acid sequences of suitable Cas9 domains and Cas9 fragments are provided herein, and additional suitable sequences of Cas9 domains and fragments will be apparent to those of skill in the art.

[0027] In some embodiments, Cas9 refers to Cas9 from: *Corynebacterium ulcerans* (NCBI Refs: NC_015683.1, NC_017317.1); *Corynebacterium diphtheria* (NCBI Refs:

NC_016782.1, NC_016786.1); *Spiroplasma syrphidicola* (NCBI Ref: NC_021284.1); *Prevotella intermedia* (NCBI Ref: NC_017861.1); *Spiroplasma taiwanense* (NCBI Ref: NC_021846.1); *Streptococcus iniae* (NCBI Ref: NC_021314.1); *Belliella baltica* (NCBI Ref: NC_018010.1); *Psychroflexus torquisI* (NCBI Ref: NC_018721.1); *Streptococcus thermophilus* (NCBI Ref: YP_820832.1); *Listeria innocua* (NCBI Ref: NP_472073.1); *Campylobacter jejuni* (NCBI Ref: YP_002344900.1); or *Neisseria. meningitidis* (NCBI Ref: YP_002342100.1).

[0028] 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.

[0029] The term "effective amount," as used herein, refers to an amount of a biologically active agent 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 target site specifically bound and cleaved by the nuclease. In some embodiments, an effective amount of a fusion protein provided herein, e.g., of a fusion protein comprising a nuclease-inactive Cas9 domain and a nucleic acid-editing domain (e.g., a deaminase domain) may refer to the amount of the fusion protein that is sufficient to induce editing of a target site specifically bound and edited by the fusion protein. As will be appreciated by the skilled artisan, the effective amount of an agent, *e.g.*, a fusion protein, a nuclease, a deaminase, a recombinase, a hybrid 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, e.g., on the specific allele, genome, or target site to be edited, on the cell or tissue being targeted, and on the agent being used.

[0030] The term "linker," as used herein, refers to a chemical group or a molecule linking two molecules or moieties, *e.g.*, two domains of a fusion protein, such as, for example, a nuclease-inactive Cas9 domain and a nucleic acid-editing domain (e.g., a deaminase domain). In some embodiments, a linker joins a gRNA binding domain of an RNA-programmable nuclease, including a Cas9 nuclease domain, and the catalytic domain of anucleic-acid editing protein. In some embodiments, a linker joins a dCas9 and a nucleic-acid editing protein. 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 or a plurality of amino acids (*e.g.*, a peptide or protein). In some embodiments, the linker is an organic molecule, group,

polymer, or chemical moiety. In some embodiments, the linker is 5-100 amino acids in length, for example, 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, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, or 150-200 amino acids in length. Longer or shorter linkers are also contemplated.

[0031] 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. Various methods for making the amino acid substitutions (mutations) provided herein are well known in the art, and are provided by, for example, Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4^{th} ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)).

[0032] 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, "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, *e.g.,* 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).

[0033] 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.

[0034] 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) and a nucleic acid cleavage domain or a catalytic domain of a nucleic-acid editing protein. In some embodiments, a protein comprises a proteinaceous part, *e.g.*, an amino acid sequence constituting a nucleic acid binding domain, and an organic compound, *e.g.*, a compound that can act as a nucleic acid cleavage agent. In some embodiments, a protein is in a complex with, or is in association with, a nucleic acid, *e.g.*, 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.

[0035] 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 RNA 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 a complex 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 single molecules or as a complex of two or more molecules. Typically, gRNAs that exist as single RNA species comprise two domains: (1) a domain that shares homology to a target nucleic acid (*e.g.*, and directs 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. Other examples of gRNAs (*e.g.*, those including domain 2) can be found in U.S. Provisional Patent Application, U.S.S.N. 61/874,682, filed September 6, 2013, entitled "Switchable Cas9 Nucleases And Uses Thereof," and U.S. Provisional Patent Application, U.S.S.N. 61/874,746, filed September 6, 2013, entitled "Delivery System For Functional Nucleases," the entire contents of each are hereby incorporated by reference in their entirety.

In some embodiments, a gRNA comprises two or more of domains (1) and (2), and may be referred to as an "extended gRNA." For example, an extended gRNA will, *e.g.*, bind two or more Cas9 proteins and bind a target nucleic acid at two or more distinct regions, as described herein. The gRNA comprises a nucleotide sequence that complements a target site, which mediates binding of the nuclease/RNA complex to said target site, providing the sequence specificity of the nuclease:RNA complex. 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., White J., Yuan X., Clifton S.W., Roe B.A., 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 C.M., 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.

[0036] Because RNA-programmable nucleases (*e.g.*, Cas9) use RNA:DNA hybridization to target DNA cleavage sites, these proteins are able to be targeted, in principle, to 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, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823 (2013); Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826 (2013); Hwang, W.Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature biotechnology* **31**, 227-229 (2013); Jinek, M. *et al.* RNA-programmed genome editing in human cells. *eLife* **2**, e00471 (2013); Dicarlo, J.E. *et al.* Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic acids research* (2013); Jiang, W. *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).

[0037] The term "subject," as used herein, refers to an individual organism, for example, an individual mammal. 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.

[0038] The term "target site" refers to a sequence within a nucleic acid molecule that is deaminated by a deaminase or a fusion protein comprising a deaminase, (*e.g.*, a dCas9-deaminase fusion protein provided herein).

[0039] 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 other 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.

DETAILED DESCRIPTION OF CERTAIN EMBODIMENTS OF THE INVENTION

[0040] Some aspects of this disclosure provide fusion proteins that comprise a Cas9 domain that binds to a guide RNA (also referred to as gRNA or sgRNA), which, in turn, binds a target nucleic acid sequence via strand hybridization; and a DNA-editing domain, for example, a deaminase domain that can deaminate a nucleobase, such as, for example, cytidine. The deamination of a nucleobase by a deaminase can lead to a point mutation at the respective residue, which is referred to herein as nucleic acid editing. Fusion proteins comprising a Cas9 variant or domain and a DNA editing domain can thus be used for the targeted editing of nucleic acid sequences. Such fusion proteins are useful for targeted editing of DNA *in vitro*, *e.g.*, for the generation of mutant cells or animals; for the introduction of targeted mutations, *e.g.*, for the correction of genetic defects in cells *ex vivo*,

e.g., in cells obtained from a subject that are subsequently re-introduced into the same or another subject; and for the introduction of targeted mutations, *e.g.*, the correction of genetic defects or the introduction of deactivating mutations in disease-associated genes in a subject. Typically, the Cas9 domain of the fusion proteins described herein does not have any nuclease activity but instead is a Cas9 fragment or a dCas9 protein or domain. Methods for the use of Cas9 fusion proteins as described herein are also provided.

[0041] Non-limiting, exemplary nuclease-inactive Cas9 domains are provided herein. One exemplary suitable nuclease-inactive Cas9 domain is the D10A/H840A Cas9 domain mutant:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR
KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENL
IAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKF
IKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVD
QELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD
WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISE
FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ
SITGLYETRIDLSQLGGD (SEQ ID NO: 37; see, *e.g.*, Qi *et al.*, Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell. 2013; 152(5):1173-83, the entire contents of which are incorporated herein by reference).

[0042] Additional suitable nuclease-inactive Cas9 domains will be apparent to those of skill in the art based on this disclosure. Such additional exemplary suitable nuclease-inactive Cas9 domains include, but are not limited to, D10A, D10A/D839A/H840A, and D10A/D839A/H840A/N863A mutant domains (See, *e.g.*, Prashant *et al.*, CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nature Biotechnology. 2013; 31(9): 833-838, the entire contents of which are incorporated herein by reference).

Fusion proteins between Cas9 and nucleic acid editing enzymes or domains

[0043] Some aspects of this disclosure provide fusion proteins comprising (i) a nuclease-inactive Cas9 enzyme or domain; and (ii) a nucleic acid-editing enzyme or domain. In some embodiments, the nucleic acid-editing enzyme or domain is a DNA-editing enzyme

or domain. In some embodiments, the nucleic acid-editing enzyme possesses deaminase activity. In some embodiments, the nucleic acid-editing enzyme or domain comprises or is a deaminase domain. In some embodiments, the deaminase is a cytidine deaminase. In some embodiments, the deaminase is an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase. In some embodiments, the deaminase is an APOBEC1 family deaminase. In some embodiments, the deaminase is an activation-induced cytidine deaminase (AID). In some embodiments, the deaminase is an ACF1/ASE deaminase. In some embodiments, the deaminase is an adenosine deaminase. In some embodiments, the deaminase is an ADAT family deaminase. Some nucleic-acid editing enzymes and domains as well as Cas9 fusion proteins including such enzymes or domains are described in detail herein. Additional suitable nucleic acid-editing enzymes or domains will be apparent to the skilled artisan based on this disclosure.

[0044] The instant disclosure provides Cas9:nucleic acid-editing enzyme/domain fusion proteins of various configurations. In some embodiments, the nucleic acid-editing enzyme or domain is fused to the N-terminus of the Cas9 domain. In some embodiments, the nucleic acid-editing enzyme or domain is fused to the C-terminus of the Cas9 domain. In some embodiments, the Cas9 domain and the nucleic acid-editing-editing enzyme or domain are fused via a linker. In some embodiments, the linker comprises a $(GGGGS)_n$ (SEQ ID NO: 91), a $(G)_n$, an $(EAAAK)_n$ (SEQ ID NO: 5), a $(GGS)_n$, an SGSETPGTSESATPES (SEQ ID NO: 93) motif (see, *e.g.*, Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* **2014;** 32(6): 577-82; the entire contents are incorporated herein by reference), or an $(XP)_n$ motif, or a combination of any of these, wherein n is independently an integer between 1 and 30. In some embodiments, n is independently 1, 2, 3, 4, 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, or 30, or, if more than one linker or more than one linker motif is present, any combination thereof. Additional suitable linker motifs and linker configurations will be apparent to those of skill in the art. In some embodiments, suitable linker motifs and configurations include those described in Chen *et al.*, Fusion protein linkers: property, design and functionality. Adv Drug Deliv Rev. 2013; 65(10):1357-69, the entire contents of which are incorporated herein by reference. Additional suitable linker sequences will be apparent to those of skill in the art based on the instant disclosure.

[0045] In some embodiments, the general architecture of exemplary Cas9 fusion proteins provided herein comprises the structure:

$[NH_2]$-[nucleic acid-editing enzyme or domain]-[Cas9]-[COOH] or

$[NH_2]$-[Cas9]-[nucleic acid-editing enzyme or domain]-[COOH],

wherein NH_2 is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein..

[0046] Additional features may be present, for example, one or more linker sequences between the NLS and the rest of the fusion protein and/or between the nucleic acid-editing enzyme or domain and the Cas9. Other exemplary features that may be present are localization sequences, such as nuclear localization sequences, 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 include, but are not limited to, biotin carboxylase carrier protein (BCCP) tags, myc-tags , calmodulin-tags, FLAG-tags, 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.

[0047] In some embodiments, the nucleic acid-editing enzyme or domain is a deaminase. For example, in some embodiments, the general architecture of exemplary Cas9 fusion proteins with a deaminase enzyme or domain comprises the structure:

$[NH_2]$-[NLS]-[Cas9]-[deaminase]-[COOH],

$[NH_2]$-[NLS]-[deaminase]-[Cas9]- -[COOH],

$[NH_2]$-[Cas9]-[deaminase]-[COOH], or

$[NH_2]$-[deaminase]-[Cas9]-[COOH]

wherein NLS is a nuclear localization signal, 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 Cas9 and the deaminase. In some embodiments, the NLS is located C-terminal of the deaminase and/or the Cas9 domain. In some embodiments, the NLS is located between the deaminase and the Cas9 domain. Additional features, such as sequence tags, may also be present

[0048] One exemplary suitable type of nucleic acid-editing enzymes and domains are cytosine deaminases, for example, of the APOBEC family. The apolipoprotein B mRNA-editing complex (APOBEC) family of cytosine deaminase enzymes encompasses eleven proteins that serve to initiate mutagenesis in a controlled and beneficial manner.[29] One family member, activation-induced cytidine deaminase (AID), is responsible for the maturation of antibodies by converting cytosines in ssDNA to uracils in a transcription-dependent, strand-biased fashion.[30] The apolipoprotein B editing complex 3 (APOBEC3) enzyme provides protection to human cells against a certain HIV-1 strain via the deamination of cytosines in reverse-transcribed viral ssDNA.[31] These proteins all require a Zn^{2+}-coordinating motif (His-X-Glu-X_{23-26}-Pro-Cys-X_{2-4}-Cys) and bound water molecule for catalytic activity. The Glu residue acts to activate the water molecule to a zinc hydroxide for nucleophilic attack in the deamination reaction. Each family member preferentially deaminates at its own particular "hotspot", ranging from WRC (W is A or T, R is A or G) for hAID, to TTC for hAPOBEC3F.[32] A recent crystal structure of the catalytic domain of APOBEC3G (Figure 2) revealed a secondary structure comprised of a five-stranded β-sheet core flanked by six α-helices, which is believed to be conserved across the entire family.[33] The active center loops have been shown to be responsible for both ssDNA binding and in determining "hotspot" identity.[34] Overexpression of these enzymes has been linked to genomic instability and cancer, thus highlighting the importance of sequence-specific targeting.[35]

[0049] Another exemplary suitable type of nucleic acid-editing enzymes and domains are adenosine deaminases. For example, an ADAT family adenosine deaminase can be fused to a Cas9 domain, *e.g.*, a nuclease-inactive Cas9 domain, thus yielding a Cas9-ADAT fusion protein.

[0050] Some aspects of this disclosure provide a systematic series of fusions between Cas9 and deaminase enzymes, *e.g.*, cytosine deaminase enzymes such as APOBEC enzymes, or adenosine deaminase enzymes such as ADAT enzymes, that has been generated in order to direct the enzymatic activities of these deaminases to a specific site in genomic DNA. The advantages of using Cas9 as the recognition agent are twofold: (1) the sequence specificity of Cas9 can be easily altered by simply changing the sgRNA sequence; and (2) Cas9 binds to its target sequence by denaturing the dsDNA, resulting in a stretch of DNA that is single-stranded and therefore a viable substrate for the deaminase. Successful fusion proteins have been generated with human and mouse deaminase domains, *e.g.*, AID domains. A variety of other fusion proteins between the catalytic domains of human and mouse AID and Cas9 are also contemplated. It will be understood that other catalytic domains, or catalytic domains

from other deaminases, can also be used to generate fusion proteins with Cas9, and that the disclosure is not limited in this regard.

[0051] In some embodiments, fusion proteins of Cas9 and AID are provided. In an effort to engineer Cas9 fusion proteins to increase mutation rates in ssDNA, both mouse and human AID were tethered to gene V of filamentous phage (a nonspecific ssDNA binding protein). The resulting fusion proteins exhibited enhanced mutagenic activities compared to the wild type enzymes in a cell-based assay. This work demonstrates that the enzymatic activity of these proteins is maintained in and can be successfully targeted to genetic sequences with fusion proteins.[36]

[0052] While several crystal structures of Cas9 (and even Cas9 in complex with its sgRNA and target DNA) have been reported, (see, *e.g.*, Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S, Kaplan M, Iavarone AT, Charpentier E, Nogales E, Doudna JA. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science.* **2014;** 343(6176):1247997. PMID: 24505130; and Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O. Crystal structure of Cas9 in c omplex with guide RNA and target DNA. *Cell.* **2014;** 156(5):935-49. PMID: 24529477, the entire contents of each of which are incorporated herein by reference), the portion of DNA that is single stranded in the Cas9-DNA complex is unknown (the size of the Cas9-DNA bubble). However, it has been shown in a dCas9 system with a sgRNA specifically designed for the complex to interfere with transcription that transcriptional interference only occurs when the sgRNA binds to the non-template strand. This result suggests that certain portions of the DNA in the DNA-Cas9 complex are unguarded by Cas9, and could potentially be targeted by a deaminase in the fusion protein (see Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell.* **2013;** 152(5):1173-83. PMID: 23452860, the entire contents of which are incorporated herein by reference). Further supporting this notion, footprinting experiments with exonuclease III and nuclease P1 (which only acts on ssDNA as a substrate) have revealed that at least 26 bases on the non-template strand are susceptible to digestion by these enzymes (see Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S, Kaplan M, Iavarone AT, Charpentier E, Nogales E, Doudna JA. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science.* **2014;** 343(6176):1247997. PMID: 24505130). It has also been reported that in certain cases, Cas9 induces single base-substitution mutations in this susceptible stretch of DNA at frequencies

as high as 15% (see Tsai SQ, Wyvekens N, Khayter C, Foden JA, Thapar V, Reyon D, Goodwin MJ, Aryee MJ, Joung JK. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol.* **2014;** 32(6):569-76. PMID: 24770325, the entire contents of which are incorporated herein by reference). While the mechanism of introduction of these mutations is unknown, in all cases, the base that is mutated is a cytosine, which could possibly indicate the involvement of a cytosine deaminase enzyme. Taken together, these data are clearly consistent with a portion of the target DNA being single stranded and susceptible to other enzymes. It has been shown in a dCas9 system with a sgRNA specifically designed for the complex to interfere with transcription that transcriptional interference only occurs when the sgRNA binds to the non-template strand. This result suggests that certain portions of the DNA in the DNA-Cas9 complex are unguarded by Cas9, and could potentially be targeted by AID in the fusion protein.[16] Accordingly, both N-terminal and C-terminal fusions of Cas9 with a deaminase domain are useful according to aspects of this disclosure.

[0053] In some embodiments, the deaminase domain and the Cas9 domain are fused to each other via a linker. Various linker lengths and flexibilities between the deaminase domain (*e.g.*, AID) and the Cas9 domain can be employed (*e.g.*, ranging from very flexible linkers of the form $(GGGGS)_n$ (SEQ ID NO: 91), $(GGS)_n$, and $(G)_n$ to more rigid linkers of the form $(EAAAK)_n$ (SEQ ID NO: 5), SGSETPGTSESATPES (SEQ ID NO: 93) (see, *e.g.*, Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* **2014;** 32(6): 577-82; the entire contents are incorporated herein by reference) and $(XP)_n$)[37] in order to achieve the optimal length for deaminase activity for the specific application.

[0054] Some exemplary suitable nucleic-acid editing enzymes and domains, *e.g.*, deaminases and deaminase domains, that can be fused to Cas9 domains 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).

[0055] Human AID:

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGC HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTAR LYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHEN SVRLSRQLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 6)

(underline: nuclear localization signal; double underline: nuclear export signal)

[0056] Mouse AID:

MDSLLMKQKKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSCSLDFGHLRNKSGC HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPNLSLRIFTAR LYFCEDRKAEPEGLRRLHRAGVQIGIMTFKDYFYCWNTFVENRERTFKAWEGLHEN SVRLTRQLRRILLPLYEVDDLRDAFRMLGF (SEQ ID NO: 7)

(underline: nuclear localization signal; double underline: nuclear export signal)

[0057] Dog AID:

MDSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGC HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAAR LYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENREKTFKAWEGLHEN SVRLSRQLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 8)

(underline: nuclear localization signal; double underline: nuclear export signal)

[0058] Bovine AID:

MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGC HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTAR LYFCDKERKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHE NSVRLSRQLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 9)

(underline: nuclear localization signal; double underline: nuclear export signal)

[0059] Mouse APOBEC-3:

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPV SLHHGVFKNKDNI*HAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFEC*AEQIVRFL ATHHNLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDN GGRRFRPWKRLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETRFCVEG RRMDPLSEEEFYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGK *QHAEILFLDKIRSMELSQVTITCYLTWSPCPNC*AWQLAAFKRDRPDLILHIYTSRLYFHW KRPFQKGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRL RRIKESWGLQDLVNDFGNLQLGPPMS (SEQ ID NO: 10)

(italic: nucleic acid editing domain)

[0060] Rat APOBEC-3:

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLRYAIDRKDTFLCYEVTRKDCDSPVS LHHGVFKNKDNI*HAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFEC*AEQVLRFLA THHNLSLDIFSSRLYNIRDPENQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGG RRFRPWKKLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETRFCVERRR VHLLSEEEFYSQFYNQRVKHLCYYHGVKPYLCYQLEQFNGQAPLKGCLLSEKGKQ*H AEILFLDKIRSMELSQVIITCYLTWSPCPNC*AWQLAAFKRDRPDLILHIYTSRLYFHWKR PFQKGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLHR IKESWGLQDLVNDFGNLQLGPPMS (SEQ ID NO: 11)

(italic: nucleic acid editing domain)

[0061] Rhesus macaque APOBEC-3G:

MVEPMDPRTFVSNFNNRPILSGLNTVWLCCEVKTKDPSGPPLDAKIFQGKVYSKAKY *HPEMRFLRWFHKWRQLHHDQEYKVTWYVSWSPCTRC*ANSVATFLAKDPKVTLTIFVA RLYYFWKPDYQQALRILCQKRGGPHATMKIMNYNEFQDCWNKFVDGRGKPFKPRN NLPKHYTLLQATLGELLRHLMDPGTFTSNFNNKPWVSGQHETYLCYKVERLHNDT WVPLNQHRGFLRNQAPNIHGFPKGR*HAELCFLDLIPFWKLDGQQYRVTCFTSWSPCFS C*AQEMAKFISNNEHVSLCIFAARIYDDQGRYQEGLRALHRDGAKIAMMNYSEFEYC WDTFVDRQGRPFQPWDGLDEHSQALSGRLRAI (SEQ ID NO: 12)

(italic: nucleic acid editing domain; underline: cytoplasmic localization signal)

[0062] Chimpanzee APOBEC-3G:

MKPHFRNPVERMYQDTFSDNFYNRPILSHRNTVWLCYEVKTKGPSRPPLDAKIFRGQ VYSKLKY*HPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKC*TRDVATFLAEDPKV TLTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRE LFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTSNFNNELWVRGRHETYLCYEVERL HNDTWVLLNQRRGFLCNQAPHKHGFLEGR*HAELCFLDVIPFWKLDLHQDYRVTCFTS WSPCFSC*AQEMAKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLAKAGAKISIMTYSE FKHCWDTFVDHQGCPFQPWDGLEEHSQALSGRLRAILQNQGN (SEQ ID NO: 13)

(italic: nucleic acid editing domain; underline: cytoplasmic localization signal)

[0063] Green monkey APOBEC-3G:

MNPQIRNMVEQMEPDIFVYYFNNRPILSGRNTVWLCYEVKTKDPSGPPLDANIFQGK LYPEAKD*HPEMKFLHWFRKWRQLHRDQEYEVTWYVSWSPCTRC*ANSVATFLAEDPKV TLTIFVARLYYFWKPDYQQALRILCQERGGPHATMKIMNYNEFQHCWNEFVDGQG KPFKPRKNLPKHYTLLHATLGELLRHVMDPGTFTSNFNNKPWVSGQRETYLCYKVE RSHNDTWVLLNQHRGFLRNQAPDRHGFPKGR*HAELCFLDLIPFWKLDDQQYRVTCFT*

*SWSPCFSC*AQKMAKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLHRDGAKIAVMNY SEFEYCWDTFVDRQGRPFQPWDGLDEHSQALSGRLRAI (SEQ ID NO: 14)

(italic: nucleic acid editing domain; underline: cytoplasmic localization signal)

[0064] Human APOBEC-3G:

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQ VYSELKY*HPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKC*TRDMATFLAEDPKV TLTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRE LFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERM HNDTWVLLNQRRGFLCNQAPHKHGFLEGR*HAELCFLDVIPFWKLDLDQDYRVTCFTS WSPCFSC*AQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSE FKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN (SEQ ID NO: 15)

(italic: nucleic acid editing domain; underline: cytoplasmic localization signal)

[0065] Human APOBEC-3F:

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQ VYSQPEH*HAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDC*VAKLAEFLAEHPNVTL TISAARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYSEGQPFMPW YKFDDNYAFLHRTLKEILRNPMEAMYPHIFYFHFKNLRKAYGRNESWLCFTMEVVK HHSPVSWKRGVFRNQVDPETHC*HAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPEC*A GEVAEFLARHSNVNLTIFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCW ENFVYNDDEPFKPWKGLKYNFLFLDSKLQEILE (SEQ ID NO: 16)

(italic: nucleic acid editing domain)

[0066] Human APOBEC-3B:

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFR GQVYFKPQY*HAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDC*VAKLAEFLSEHPN VTLTISAARLYYYWERDYRRALCRLSQAGARVTIMDYEEFAYCWENFVYNEGQQF MPWYKFDENYAFLHRTLKEILRYLMDPDTFTFNFNNDPLVLRRRQTYLCYEVERLD NGTWVLMDQHMGFLCNEAKNLLCGFY*GRHAELRFLDLVPSLQLDPAQIYRVTWFISWS PCFSWGC*AGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTY DEFEYCWDTFVYRQGCPFQPWDGLEEHSQALSGRLRAILQNQGN (SEQ ID NO: 17)

(italic: nucleic acid editing domain)

[0067] Human APOBEC-3C:

MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVF RNQVDSETH*CHAERCFLSWFCDDILSPNTKYQVTWYTSWSPCPDC*AGEVAEFLARHSN

VNLTIFTARLYYFQYPCYQEGLRSLSQEGVAVEIMDYEDFKYCWENFVYNDNEPFKP WKGLKTNFRLLKRRLRESLQ (SEQ ID NO: 18)

(italic: nucleic acid editing domain)

[0068] Human APOBEC-3A:

MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLH NQAKNLLCGFYGR*HAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGC*AGEVRAFLQ ENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGC PFQPWDGLDEHSQALSGRLRAILQNQGN (SEQ ID NO: 19)

(italic: nucleic acid editing domain)

[0069] Human APOBEC-3H:

MALLTAETFRLQFNNKRRLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKC*HAEI CFINEIKSMGLDETQCYQVTCYLTWSPCSSC*AWELVDFIKAHDHLNLGIFASRLYYHWC KPQQKGLRLLCGSQVPVEVMGFPKFADCWENFVDHEKPLSFNPYKMLEELDKNSRA IKRRLERIKIPGVRAQGRYMDILCDAEV (SEQ ID NO: 20)

(italic: nucleic acid editing domain)

[0070] Human APOBEC-3D:

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFR GPVLPKRQSNHRQEVYFRFEN*HAEMCFLSWFCGNRLPANRRFQITWFVSWNPCLPC*VV KVTKFLAEHPNVTLTISAARLYYYRDRDWRWVLLRLHKAGARVKIMDYEDFAYCW ENFVCNEGQPFMPWYKFDDNYASLHRTLKEILRNPMEAMYPHIFYFHFKNLLKACG RNESWLCFTMEVTKHHSAVFRKRGVFRNQVDPETHC*HAERCFLSWFCDDILSPNTNY EVTWYTSWSPCPEC*AGEVAEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGAS VKIMGYKDFVSCWKNFVYSDDEPFKPWKGLQTNFRLLKRRLREILQ (SEQ ID NO: 21)

(italic: nucleic acid editing domain)

[0071] Human APOBEC-1:

MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKN TTNHVEVNFIKKFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYV ARLFWHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQY PPLWMMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPPHILLATGLI HPSVAWR (SEQ ID NO: 22)

[0072] Mouse APOBEC-1:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSVWRHTSQN TSNHVEVNFLEKFTTERYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPYVTLFIYIA

RLYHHTDQRNRQGLRDLISSGVTIQIMTEQEYCYCWRNFVNYPPSNEAYWPRYPHL WVKLYVLELYCIILGLPPCLKILRRKQPQLTFFTITLQTCHYQRIPPHLLWATGLK (SEQ ID NO: 23)

[0073] Rat APOBEC-1:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNT NKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIAR LYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLW VRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK (SEQ ID NO: 24)

[0074] Human ADAT-2:

MEAKAAPKPAASGACSVSAEETEKWMEEAMHMAKEALENTEVPVGCLMVYNNEV VGKGRNEVNQTKNATRHAEMVAIDQVLDWCRQSGKSPSEVFEHTVLYVTVEPCIM CAAALRLMKIPLVVYGCQNERFGGCGSVLNIASADLPNTGRPFQCIPGYRAEEAVEM LKTFYKQENPNAPKSKVRKKECQKS (SEQ ID NO: 25)

[0075] Mouse ADAT-2:

MEEKVESTTTPDGPCVVSVQETEKWMEEAMRMAKEALENIEVPVGCLMVYNNEVV GKGRNEVNQTKNATRHAEMVAIDQVLDWCHQHGQSPSTVFEHTVLYVTVEPCIMC AAALRLMKIPLVVYGCQNERFGGCGSVLNIASADLPNTGRPFQCIPGYRAEEAVELL KTFYKQENPNAPKSKVRKKDCQKS (SEQ ID NO: 26)

[0076] Mouse ADAT-1:

MWTADEIAQLCYAHYNVRLPKQGKPEPNREWTLLAAVVKIQASANQACDIPEKEVQ VTKEVV*SMGTGTKCIGQSKMRESGDILNDSHAEIIARRSFQRYLLHQLHLAAVLKEDSIFV PGTQRGLWRLRPDLSFVFFSSHTPCGDASIIPMLEFEEQPCCPVIRSWANNSPVQETENLE DSKDKRNCEDPASPVAKKMRLGTPARSLSNCVAHHGTQESGPVKPDVSSSDLTKEEPDAA NGIASGSFRVVDVYRTGAKCVPGETGDLREPGAAYHQVGLLRVKPGRGDRTCSMSCSDK MARWNVLGCQGALLMHFLEKPIYLSAVVIGKCPYSQEAMRRALTGRCEETLVLPRGFGVQ ELEIQQSGLLFEQSRCAVHRKRGDSPGRLVPCGAAISWSAVPQQPLDVTANGFPQGTTKK EIGSPRARSRISKVELFRSFQKLLSSIADDEQPDSIRVTKKLDTYQEYKDAASAYQEAWGAL RRIQPFASWIRNPPDYHQF*K (SEQ ID NO: 27) (italic: nucleic acid editing domain)

[0077] Human ADAT-1:

MWTADEIAQLCYEHYGIRLPKKGKPEPNHEWTLLAAVVKIQSPADKACDTPDKPVQ VTKEVV*SMGTGTKCIGQSKMRKNGDILNDSHAEVIARRSFQRYLLHQLQLAATLKEDSIF VPGTQKGVWKLRRDLIFVFFSSHTPCGDASIIPMLEFEDQPCCPVFRNWAHNSSVEASSNL EAPGNERKCEDPDSPVTKKMRLEPGTAAREVTNGAAHHQSFGKQKSGPISPGIHSCDLTV*

*EGLATVTRIAPGSAKVIDVYRTGAKCVPGEAGDSGKPGAAFHQVGLLRVKPGRGDRTRSM SCSDKMARWNVLGCQGALLMHLLEEPIYLSAVVIGKCPYSQEAMQRALIGRCQNVSALPK GFGVQELKILQSDLLFEQSRSAVQAKRADSPGRLVPCGAAISWSAVPEQPLDVTANGFPQ GTTKKTIGSLQARSQISKVELFRSFQKLLSRIARDKWPHSLRVQKLDTYQEYKEAASSYQEA WSTLRKQVFGSWIRNPPDYHQF*K (SEQ ID NO: 28) (italic: nucleic acid editing domain)

[0078] In some embodiments, fusion proteins as provided herein comprise the full-length amino acid of a nucleic acid-editing enzyme, *e.g.*, one of the sequences provided above. In other embodiments, however, fusion proteins as provided herein do not comprise a full-length sequence of a nucleic acid-editing enzyme, but only a fragment thereof. For example, in some embodiments, a fusion protein provided herein comprises a Cas9 domain and a fragment of a nucleic acid-editing enzyme, *e.g.*, wherein the fragment comprises a nucleic acid-editing domain. Exemplary amino acid sequences of nucleic acid-editing domains are shown in the sequences above as italicized letters, and additional suitable sequences of such domains will be apparent to those of skill in the art.

[0079] Additional suitable nucleic-acid editing enzyme sequences, *e.g.*, deaminase enzyme and domain sequences, that can be used according to aspects of this invention, *e.g.*, that can be fused to a nuclease-inactive Cas9 domain, will be apparent to those of skill in the art based on this disclosure. In some embodiments, such additional enzyme sequences include deaminase enzyme or deaminase domain sequences that are at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% similar to the sequences provided herein. Additional suitable Cas9 domains, variants, and sequences will also be apparent to those of skill in the art. Examples of such additional suitable Cas9 domains include, but are not limited to, D10A, D10A/D839A/H840A, and D10A/D839A/H840A/N863A mutant domains (See, *e.g.*, Prashant *et al.*, CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nature Biotechnology. 2013; 31(9): 833-838 the entire contents of which are incorporated herein by reference).

[0080] Additional suitable strategies for generating fusion proteins comprising a Cas9 domain and a deaminase domain will be apparent to those of skill in the art based on this disclosure in combination with the general knowledge in the art. Suitable strategies for generating fusion proteins according to aspects of this disclosure using linkers or without the use of linkers will also be apparent to those of skill in the art in view of the instant disclosure and the knowledge in the art. For example, Gilbert *et al.*, CRISPR-mediated modular RNA-

guided regulation of transcription in eukaryotes. *Cell.* **2013**; 154(2):442-51, showed that C-terminal fusions of Cas9 with VP64 using 2 NLS's as a linker (SPKKKRKVEAS, SEQ ID NO: 29), can be employed for transcriptional activation. Mali *et al.*, CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol.* **2013**; 31(9):833-8, reported that C-terminal fusions with VP64 without linker can be employed for transcriptional activation. And Maeder *et al.*, CRISPR RNA-guided activation of endogenous human genes. *Nat Methods.* **2013**; 10: 977-979, reported that C-terminal fusions with VP64 using a Gly_4Ser (SEQ ID NO: 91) linker can be used as transcriptional activators. Recently, dCas9- FokI nuclease fusions have successfully been generated and exhibit improved enzymatic specificity as compared to the parental Cas9 enzyme (In Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* **2014;** 32(6): 577-82, and in Tsai SQ, Wyvekens N, Khayter C, Foden JA, Thapar V, Reyon D, Goodwin MJ, Aryee MJ, Joung JK. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol.* **2014;** 32(6):569-76. PMID: 24770325 a SGSETPGTSESATPES (SEQ ID NO: 93) or a GGGGS (SEQ ID NO: 91) linker was used in FokI-dCas9 fusion proteins, respectively).

Use of Cas9 DNA editing fusion proteins for correcting disease-associated mutations

[0081] Some embodiments provide methods for using the Cas9 DNA editing fusion proteins provided herein. 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 C residue. 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 Cas9 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.

[0082] In some embodiments, the purpose of the methods provide herein is to restore the function of a dysfunctional gene via genome editing. The Cas9 deaminase fusion proteins provided herein 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 Cas9 domain 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.

[0083] 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.[50] In fact, the A3140G mutation is present in several NCI-60 cancer cell lines, such as, for example, the HCT116, SKOV3, and T47D cell lines, which are readily available from the American Type Culture Collection (ATCC).[51]

[0084] 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 Cas9 deaminase fusion protein and an appropriately designed sgRNA targeting the fusion protein to the respective mutation site in the encoding PI3KCA gene. Control experiments can be performed where the sgRNAs 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.

[0085] 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 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.

[0086] 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 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), Gln (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*.

[0087] 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 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 Cas9 deaminase fusion protein that corrects the point mutation or introduces a deactivating mutation into the disease-associated gene. 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.

[0088] 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 are listed below. It will be understood that the numbering of the specific positions or residues in the respective sequences depends on the particular protein and numbering scheme used. Numbering might be different, *e.g.*, in precursors of a mature protein and the mature protein itself, and differences in sequences from species to species may affect numbering. One of skill in the art will be able to identify the respective residue in any homologous protein and in the respective encoding nucleic acid by methods well known in the art, *e.g.*, by sequence alignment and determination of homologous residues. 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 (mouse) or 240 (human) or a homologous residue 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 or a homologous residue, or cysteine to arginine at residue 24 or a homologous residue in the platelet membrane glycoprotein IX (T>C mutation) – see, *e.g.*, Noris *et al.*, *British Journal of Haematology.* **1997;** 97: 312-320, and Ali *et al.*, *Hematol.* **2014;** 93: 381-384; epidermolytic hyperkeratosis (EHK) – *e.g.*, leucine to proline mutation at position 160 or 161 (if counting the initiator methionine) or a homologous residue in keratin 1 (T>C mutation) – see, *e.g.*, Chipev *et al.*, *Cell.* **1992;** 70: 821-828, see also accession number P04264 in the UNIPROT database at www[dot]uniprot[dot]org; chronic obstructive pulmonary disease (COPD) – *e.g.*, leucine to proline mutation at position 54 or 55 (if counting the initiator methionine) or a homologous residue in the processed form of α_1-antitrypsin or residue 78 in the unprocessed form or a homologous residue (T>C mutation) – see, *e.g.*, Poller *et al.*, *Genomics.* **1993;** 17: 740-743, see also accession number P01011 in the UNIPROT database; Charcot-Marie-Toot disease type 4J – *e.g.*, isoleucine to threonine mutation at position 41 or a homologous residue in FIG4 (T>C mutation) – see, *e.g.*, Lenk *et al.*, PLoS Genetics. 2011; 7: e1002104; neuroblastoma (NB) – *e.g.*, leucine to proline mutation at position 197 or a homologous residue in Caspase-9 (T>C mutation) – see, *e.g.*, Kundu *et al.*, *3 Biotech.* **2013, 3:225-234**; von Willebrand disease (vWD) – *e.g.*, cysteine to arginine mutation at position 509 or a homologous residue in the processed form of von Willebrand factor, or at position 1272 or a homologous residue in the unprocessed form of von Willebrand factor (T>C mutation) – see, *e.g.*, Lavergne *et al.*, *Br. J. Haematol.* **1992**, see also accession number P04275 in the UNIPROT database**;** 82: 66-72; myotonia congenital – *e.g.*, cysteine to arginine mutation at position 277 or a homologous residue 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 or a homologous residue in the processed form of apolipoprotein AII or at position 101 or a homologous residue in the unprocessed form (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 or a homologous residue 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 or a

homologous residue in VEGFR3 tyrosine kinase (A>G mutation), see, *e.g.*, Irrthum *et al.*, *Am. J. Hum. Genet.* **2000;** 67: 295-301; familial Alzheimer's disease – *e.g.*, isoleucine to valine mutation at position 143 or a homologous residue 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 or a homologous residue 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 or a homologous residue 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 or a homologous residue in αB crystallin (A>G mutation) – see, *e.g.*, Kumar *et al.*, *J. Biol. Chem.* **1999;** 274: 24137-24141. The entire contents of all references and database entries is incorporated herein by reference.

[0089] It will be apparent to those of skill in the art that in order to target a Cas9:nucleic acid-editing enzyme/domain fusion protein as disclosed herein to a target site, e.g., a site comprising a point mutation to be edited, it is typically necessary to co-express the Cas9:nucleic acid-editing enzyme/domain fusion protein together with a guide RNA, *e.g.*, an sgRNA. As explained in more detail elsewhere herein, a guide RNA typically comprises a tracrRNA framework allowing for Cas9 binding, and a guide sequence, which confers sequence specificity to the Cas9:nucleic acid-editing enzyme/domain fusion protein. In some embodiments, the guide RNA comprises a structure 5'-[guide sequence]-guuuuagagcuagaaauagcaaguuaaaauaaaggcuaguccguuaucaacuugaaaaaguggcaccgagucggugcuuuu-3' (SEQ ID NO: 38), wherein the guide sequence comprises a sequence that is complementary to the target sequence. The guide sequence is typically 20 nucleotides long. The sequences of suitable guide RNAs for targeting Cas9:nucleic acid-editing enzyme/domain fusion proteins to specific genomic target sites will be apparent to those of skill in the art based on the instant disclosure. Such suitable guide RNA sequences typically comprise guide sequences that are complementary to a nucleic sequence within 50 nucleotides upstream or downstream of the target nucleotide to be edited. Some exemplary guide RNA sequences suitable for targeting Cas9:nucleic acid-editing enzyme/domain fusion proteins to specific target sequences are provided below.

[0090] H1047R (A3140G) polymorphism in the phosphoinositide-3-kinase catalytic alpha subunit (PI3KCA or PIK3CA) (the position of the mutated nucleotide and the respective codon are underlined):

gatgacattgcatacattcgaaagaccctagccttagataaaactgagcaagaggctttg
D D I A Y I R K T L A L D K T E Q E A L
gagtatttcatgaaacaaatgaatgatgcac**g**tcatggtggctggacaacaaaaatggat
E Y F M K Q M N D A **R** H G G W T T K M D
tggatcttccacacaattaaacagcatgcattgaactgaaagataactgagaaaatgaaa
W I F H T I K Q H A L N – K I T E K M K

(Nucleotide sequence – SEQ ID NO: 39; protein sequence – SEQ ID NO: 40).

[0091] Exemplary suitable guide sequences for targeting a Cas9:nucleic acid-editing enzyme/domain fusion proteins to the mutant A3140G residue include, without limitation: 5'-aucggaauctauuuugacuc-3' (SEQ ID NO: 41); 5'-ucggaaucuauuuugacucg-3' (SEQ ID NO: 42); 5'- cuuagauaaaacugagcaag-3' (SEQ ID NO: 43); 5'- aucuauuuugacucguucuc-3' (SEQ ID NO: 44); 5'- uaaaacugagcaagaggcuu-3' (SEQ ID NO: 45); 5'- ugguggcuggacaacaaaaa-3' (SEQ ID NO: 46); 5'- gcuggacaacaaaaauggau-3' (SEQ ID NO: 47); 5'-guguuaauuugucguacgua-3' (SEQ ID NO: 48). Additional suitable guide sequences for targeting a Cas9:nucleic acid-editing enzyme/domain fusion protein to a mutant PI3KCA sequence, to any of the additional sequences provided below, or to additional mutant sequences associated with a disease will be apparent to those of skill in the art based on the instant disclosure.

[0092] Phenylketonuria phenylalanine to serine mutation at residue 240 in phenylalanine hydroxylase gene (T>C mutation) (the position of the mutated nucleotide and the respective codon are underlined):

aatcacatttttccacttcttgaaaagtactgtggcttccatgaagataacattccccag
N H I F P L L E K Y C G F H E D N I P Q
ctggaagacgtttctcaattcctgcagacttgcactggt**c**tccgcctccgacctgtggct
L E D V S Q F L Q T C T G **S** R L R P V A
ggcctgctttcctctcgggatttcttgggtggcctggccttccgagtcttccactgcaca
G L L S S R D F L G G L A F R V F H C T

(Nucleotide sequence – SEQ ID NO: 49; protein sequence – SEQ ID NO: 50).

[0093] Bernard-Soulier syndrome (BSS) – cysteine to arginine at residue 24in the platelet membrane glycoprotein IX (T>C mutation):

atgcctgcctggggagccctgttcctgctctgggccacagcagaggccaccaaggactgc
M P A W G A L F L L W A T A E A T K D C
cccagccca**c**gtacctgccgcgcccctggaaaccatggggctgtgggtggactgcaggggc
P S P **R** T C R A L E T M G L W V D C R G

cacggactcacggccctgcctgccctgccggcccgcacccgccaccttctgctggccaac
H G L T A L P A L P A R T R H L L L A N

(Nucleotide sequence – SEQ ID NO: 51; protein sequence – SEQ ID NO: 52).

[0094] Epidermolytic hyperkeratosis (EHK) – leucine to proline mutation at residue 161 in keratin 1 (T>C mutation):

ggttatggtcctgtctgccctcctggtggcatacaagaagtcactatcaaccagagcc**C**t
G Y G P V C P P G G I Q E V T I N Q S **P**
cttcagcccctcaatgtggagattgaccctgagatccaaaaggtgaagtctcgagaaagg
L Q P L N V E I D P E I Q K V K S R E R

(Nucleotide sequence – SEQ ID NO: 53; protein sequence – SEQ ID NO: 54).

[0095] Chronic obstructive pulmonary disease (COPD) – leucine to proline mutation at residue 54 in α_1-antitrypsin (T>C mutation):

gtctccctggctgaggatccccagggagatgctgcccagaagacagatacatcccaccat
V S L A E D P Q G D A A Q K T D T S H H
gatcaggatcacccaaccttcaacaagatcacccccaacc**C**ggctgagttcgccttcagc
D Q D H P T F N K I T P N **P** A E F A F S
ctataccgccagctggcacaccagtccaacagcaccaatatcttcttctccccagtgagc
L Y R Q L A H Q S N S T N I F F S P V S

(Nucleotide sequence – SEQ ID NO: 55; protein sequence – SEQ ID NO: 56).

[0096] Chronic obstructive pulmonary disease (COPD) – leucine to proline mutation at residue 78 in α1-antichymotrypsin (T>C mutation):

gcctccgccaacgtggacttcgctttcagcctgtacaagcagttagtcctgaaggcccct
A S A N V D F A F S L Y K Q L V L K A P
gataagaatgtcatcttctccccaccgagcatctccaccgccttggccttcctgtctctg
D K N V I F S P P S I S T A L A F L S L
ggggcccataataccaccctgacagagattctcaaaggcctcaagttctacctcacggag
G A H N T T L T E I L K G L K F Y L T E

(Nucleotide sequence – SEQ ID NO: 89; protein sequence – SEQ ID NO: 90).

[0097] Neuroblastoma (NB) – leucine to proline mutation at residue 197 in Caspase-9 (T>C mutation):

ggccactgcctcattatcaacaatgtgaacttctgccgtgagtccgggctccgcacccgc
G H C L I I N N V N F C R E S G L R T R
actggctccaacatcgactgtgagaagttgcggcgtcgcttctcctcgc**C**gcatttcatg
T G S N I D C E K L R R R F S S **P** H F M
gtggaggtgaagggcgacctgactgccaagaaaatggtgctggctttgctggagctggcg
V E V K G D L T A K K M V L A L L E L A

(Nucleotide sequence – SEQ ID NO: 57; protein sequence – SEQ ID NO: 58).

[0098] Charcot-Marie-Tooth disease type 4J – isoleucine to threonine mutation at residue 41 in FIG4 (T>C mutation):

```
actagagctagatactttctagttgggagcaataatgcagaaacgaaatatcgtgtcttg
 T  R  A  R  Y  F  L  V  G  S  N  N  A  E  T  K  Y  R  V  L
aagactgatagaacagaaccaaaagatttggtcataattgatgacaggcatgtctatact
 K  T  D  R  T  E  P  K  D  L  V  I  I  D  D  R  H  V  Y  T
caacaagaagtaagggaacttcttggccgcttggatcttggaaatagaacaaagatggga
 Q  Q  E  V  R  E  L  L  G  R  L  D  L  G  N  R  T  K  M  G
```

(Nucleotide sequence – SEQ ID NO: 59; protein sequence – SEQ ID NO: 60).

[0099] von Willebrand disease (vWD) – cysteine to arginine mutation at residue 1272 in von Willebrand factor (T>C mutation):

```
acagatgccccggtgagccccaccactctgtatgtggaggacatctcggaaccgccgttg
 T  D  A  P  V  S  P  T  T  L  Y  V  E  D  I  S  E  P  P  L
cacgatttctaccgcagcaggctactggacctggtcttcctgctggatggctcctccagg
 H  D  F  Y  R  S  R  L  L  D  L  V  F  L  L  D  G  S  S  R
ctgtccgaggctgagtttgaagtgctgaaggcctttgtggtggacatgatggagcggctg
 L  S  E  A  E  F  E  V  L  K  A  F  V  V  D  M  M  E  R  L
```

(Nucleotide sequence – SEQ ID NO: 61; protein sequence – SEQ ID NO: 62).

[00100] Myotonia congenital – cysteine to arginine mutation at position 277 in the muscle chloride channel gene CLCN1 (T>C mutation):

```
atctgtgctgctgtcctcagcaaattcatgtctgtgttctgcggggtatatgagcagcca
 I  C  A  A  V  L  S  K  F  M  S  V  F  C  G  V  Y  E  Q  P
tactactactctgatatcctgacggtgggctgtgctgtgggagtcggccgttgttttggg
 Y  Y  Y  S  D  I  L  T  V  G  C  A  V  G  V  G  R  C  F  G
acaccacttggaggagtgctatttagcatcgaggtcacctccacctactttgctgttcgg
 T  P  L  G  G  V  L  F  S  I  E  V  T  S  T  Y  F  A  V  R
```

(Nucleotide sequence – SEQ ID NO: 63; protein sequence – SEQ ID NO: 64).

[00101] Hereditary renal amyloidosis – stop codon to arginine mutation at residue 111 in apolipoprotein AII (T>C mutation):

```
tactttgaaaagtcaaaggagcagctgacacccctgatcaagaaggctggaacggaactg
 Y  F  E  K  S  K  E  Q  L  T  P  L  I  K  K  A  G  T  E  L
gttaacttcttgagctatttcgtggaacttggaacacagcctgccacccagcgaagtgtc
 V  N  F  L  S  Y  F  V  E  L  G  T  Q  P  A  T  Q  R  S  V
cagcaccattgtcttccaaccccagctggcctctagaacacccactggccagtcctagag
 Q  H  H  C  L  P  T  P  A  G  L  -  N  T  H  W  P  V  L  E
```

(Nucleotide sequence – SEQ ID NO: 65; protein sequence – SEQ ID NO: 66).

[00102] Dilated cardiomyopathy (DCM) – tryptophan to Arginine mutation at position 148 in the FOXD4 gene (T>C mutation):

```
ccgcacaagcgcctcacgctcagcggcatctgcgccttcattagtgaccgcttcccctac
 P  H  K  R  L  T  L  S  G  I  C  A  F  I  S  D  R  F  P  Y
```

taccgccgcaagttccccgcc**c**ggcagaacagcatccgccacaacctctcgctgaacgac
Y R R K F P A **R** Q N S I R H N L S L N D
tgcttcgtcaagatcccccgcgagccgggccgcccaggcaagggcaactactggagcctg
C F V K I P R E P G R P G K G N Y W S L

(Nucleotide sequence – SEQ ID NO: 67; protein sequence – SEQ ID NO: 68).

[00103] Hereditary lymphedema – histidine to arginine mutation at residue 1035 in VEGFR3 tyrosine kinase (A>G mutation):

gctgaggacctgtggctgagcccgctgaccatggaagatcttgtctgctacagcttccag
A E D L W L S P L T M E D L V C Y S F Q
gtggccagagggatggagttcctggcttcccgaaagtgcatcc**g**cagagacctggctgct
V A R G M E F L A S R K C I **R** R D L A A
cggaacattctgctgtcggaaagcgacgtggtgaagatctgtgactttggccttgcccgg
R N I L L S E S D V V K I C D F G L A R

(Nucleotide sequence – SEQ ID NO: 69; protein sequence – SEQ ID NO: 70).

[00104] Familial Alzheimer's disease – isoleucine to valine mutation at residue 143 in presenilin1 (A>G mutation):

gataccgagactgtgggccagagagccctgcactcaattctgaatgctgccatcatgatc
D T E T V G Q R A L H S I L N A A I M I
agtgtc**g**ttgttgtcatgactatcctcctggtggttctgtataaatacaggtgctataag
S V **V** V V M T I L L V V L Y K Y R C Y K
gtcatccatgcctggcttattatatcatctctattgttgctgttctttttttcattcatt
V I H A W L I I S S L L L L F F F S F I

(Nucleotide sequence – SEQ ID NO: 71; protein sequence – SEQ ID NO: 72).

[00105] Prion disease – methionine to valine mutation at residue 129 in prion protein (A>G mutation):

aagccgagtaagccaaaaaccaacatgaagcacatggctggtgctgcagcagctggggca
K P S K P K T N M K H M A G A A A A G A
gtggtgggggccttggcggctac**g**tgctgggaagtgccatgagcaggcccatcatacat
V V G G L G G Y **V** L G S A M S R P I I H
ttcggcagtgactatgaggaccgttactatcgtgaaaacatgcaccgttaccccaaccaa
F G S D Y E D R Y Y R E N M H R Y P N Q

(Nucleotide sequence – SEQ ID NO: 73; protein sequence – SEQ ID NO: 74).

[00106] Chronic infantile neurologic cutaneous articular syndrome (CINCA) – Tyrosine to Cysteine mutation at residue 570 in cryopyrin (A>G mutation):

cttcccagccgagacgtgacagtccttctggaaaactatggcaaattcgaaaaggggt**g**t
L P S R D V T V L L E N Y G K F E K G **C**
ttgatttttgttgtacgtttcctctttggcctggtaaaccaggagaggacctcctacttg
L I F V V R F L F G L V N Q E R T S Y L

(Nucleotide sequence – SEQ ID NO: 75; protein sequence – SEQ ID NO: 76).

[00107] Desmin-related myopathy (DRM) – arginine to glycine mutation at residue 120 in αB crystallin (A>G mutation):

```
gtgaagcacttctccccagaggaactcaaagttaaggtgttgggagatgtgattgaggtg
 V  K  H  F  S  P  E  E  L  K  V  K  V  L  G  D  V  I  E  V
catggaaaacatgaagagcgccaggatgaacatggtttcatctccagggagttccacGgg
 H  G  K  H  E  E  R  Q  D  E  H  G  F  I  S  R  E  F  H  G
aaataccggatcccagctgatgtagaccctctcaccattacttcatccctgtcatctgat
 K  Y  R  I  P  A  D  V  D  P  L  T  I  T  S  S  L  S  S  D
```

(Nucleotide sequence – SEQ ID NO: 77; protein sequence – SEQ ID NO: 78).

[00108] Beta-thalassemia- one example is leucine to proline mutation at residue 115 in Hemoglobin B.

```
gagctgcactgtgacaagctgcacgtggatcctgagaacttcaggctcctgggcaacgtg
 E  L  H  C  D  K  L  H  V  D  P  E  N  F  R  L  L  G  N  V
ctggtctgtgtgccggcccatcactttggcaaagaattcaccccaccagtgcaggctgcc
 L  V  C  V  P  A  H  H  F  G  K  E  F  T  P  P  V  Q  A  A
tatcagaaagtggtggctggtgtggctaatgccctggcccacaagtatcactaagctcgc
 Y  Q  K  V  V  A  G  V  A  N  A  L  A  H  K  Y  H  -  A  R
```

[00109] (Nucleotide sequence – SEQ ID NO: 79; protein sequence – SEQ ID NO: 80).It is to be understood that the sequences provided above are exemplary and not meant to be limiting the scope of the instant disclosure. Additional suitable sequences of point mutations that are associated with disease and amenable to correction by Cas9:nucleic acid-editing enzyme/domain fusion proteins as well as suitable guide RNA sequences will be apparent to those of skill in the art based on this disclosure.

Reporter systems

[00110] Some aspects of this disclosure provide a reporter system that can be used for detecting deaminase activity of the fusion proteins described herein. In some embodiments, the reporter system is a luciferase-based assay in which deaminase activity leads to expression of luciferase. To minimize the impact of potential substrate promiscuity of the deaminase domain (*e.g.*, the AID domain), the number of residues that could unintentionally be targeted for deamination (*e.g.*, off-target C residues that could potentially reside on ssDNA within the reporter system) is minimized. In some embodiments, an intended target residue is be located in an ACG mutated start codon of the luciferase gene that is unable to initiate translation. Desired deaminase activity results in a ACG>AUG modification, thus enabling translation of luciferase and detection and quantification of the deaminase activity.

[00111] In some embodiments, in order to minimize single-stranded C residues, a

leader sequence is inserted between the mutated start codon and the beginning of the luciferase gene which consists of a stretch of Lys (AAA), Asn (AAT), Leu (TTA), Ile (ATT, ATA), Tyr (TAT), or Phe (TTT) residues. The resulting mutants can be tested to ensure that the leader sequence does not adversely affect luciferase expression or activity. Background luciferase activity with the mutated start codon can be determined as well.

[00112] The reporter system can be used to test many different sgRNAs, *e.g.*, in order to determine which residue(s) with respect to the target DNA sequence the respective deaminase (*e.g.*, AID enzyme) will target (Figure 3). Because the size of the Cas9-DNA bubble is not known, sgRNAs that target non-template strand can also be tested in order to assess off-target effects of a specific Cas9 deaminase fusion protein. In some embodiments, such sgRNAs are designed such that the mutated start codon will not be base-paired with the sgRNA.

[00113] Once fusion proteins that are capable of programmable site-specific C to U modifications have been identified, their activities can be further characterized. The data from the luciferase assays can, for example, be integrated into heat maps that describe which nucleotides, with respect to the sgRNA target DNA, are being targeted for deamination by a specific fusion protein. In some embodiments, the position that results in the highest activity in the luciferase assay for each fusion is considered the "target" position, while all others are considered off-target positions.

[00114] In some embodiments, Cas9 fusions with various APOBEC3 enzymes, or deaminase domains thereof, are provided. In some embodiments, Cas9 fusion proteins with other nucleic acid editing enzymes or catalytic domains are provided, including, for example, ssRNA editing enzymes, such as the cytidine deaminases APOBEC1 and ACF1/ASF, as well as the ADAT family of adenosine deaminases,[38] that can be used for ssDNA editing activity when fused to Cas9. The activity of such fusion proteins can be tested using the same reporter systems and assays described above.

[00115] In some embodiments, a reporter system is provided herein that includes a reporter gene comprising a deactivated start codon, *e.g.*, a mutation on the template strand from 3'-TAC-5' to 3'-CAC-5'. Upon successful deamination of the target C, the corresponding mRNA will be transcribed as 5'-AUG-3' instead of 5'-GUG-3', enabling the translation of the reporter gene. Suitable reporter genes will be apparent to those of skill in the art.

[00116] The description of exemplary embodiments of the reporter systems above is provided for illustration purposes only and not meant to be limiting. Additional reporter

systems, *e.g.*, variations of the exemplary systems described in detail above, are also embraced by this disclosure.

EXAMPLES

Example 1: Fusion proteins

[00117] Exemplary Cas9:deaminase fusion proteins are provided below:

[00118] *Cas9:human AID fusion (C-terminal)*

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFK
VLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDD
SFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLA
LAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKS
RRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQ
IGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL
PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTR
KSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY
VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL
GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKR
RRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG
QGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
AIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT
KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLV
SDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKV
EKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK
RMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQI
SEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD**GGGGSGGGGSGGGGS**YVVKRRDSATSFSL
DFGYLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSL
RIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHEN
SVRLSRQLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 30)

(underline: nuclear localization signal; double underline: nuclear export signal, bold: linker sequence)

[00119] *Cas9:human AID fusion (N-terminal)*

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFL
RYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPE
GLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLP**GGGG
SGGGGSGGGGS**DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL
FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE

RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP
DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF
GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGY
IDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR
RQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGA
SAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAI
VDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNE
ENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD
KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIK
KGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL
KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLT
RSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQ
LVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYH
HAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFS
KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITI
MERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPS
KYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLS
AYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG
LYETRIDLSQLGGD (SEQ ID NO: 31)

(underline: nuclear localization signal; bold: linker sequence)

[00120] *Cas9:mouse AID fusion (C-terminal)*

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFK
VLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDD
SFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLA
LAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKS
RRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQ
IGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL
PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTR
KSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY
VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL
GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKR
RRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG
QGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
AIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT
KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLV
SDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKV
EKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK
RMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQI

SEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD**GGGGSGGGGSGGGGS**YVVKRRDSATSCSL
DFGHLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPNLSL
RIFTARLYFCEDRKAEPEGLRRLHRAGVQIGIMTFKDYFYCWNTFVENRERTFKAWEGLHEN
SVRLTRQLRRILLPLYEVDDLRDAFRMLGF (SEQ ID NO: 32)
(underline: nuclear localization signal; bod: linker sequence; double underline: nuclear export signal)

[00121] *Cas9:human APOBEC-3G fusion (N-terminal)*

SPKKKRKVEASMELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAE
DPKVTLTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELF
EPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVL
LNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMA
KFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPF
QPWDGLDEHSQDLSGRLRAILQNQEN**SPKKKRKVEASSPKKKRKVEAS**KKYSIGLAIGTNSV
GWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNR
ICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPIN
ASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKL
QLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD
EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEE
LLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIEC
FDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIH
DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIV
IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDM
YVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWR
QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEN
DKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF
VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG
GFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDL
IIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ
LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNL
GAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 33)
(underline: nuclear localization signal; bold: linker (1 NLS),

[00122] *Cas9:human APOBEC-1 fusion (N-terminal)*

SPKKKRKVEASMTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIW
RSSGKNTTNHVEVNFIKKFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLV
IYVARLFWHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWM

MLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPPHILLATGLIHPSVAWR**S**
PKKKRKVEASSPKKKRKVEASDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHS
IKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEES
FLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRG
HFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQ
LPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLF
LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQI
HLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPA
FLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRL
SRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHI
ANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEE
GIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLK
DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATA
KYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVK
KTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLK
SVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGEL
QKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGD

(underline: nuclear localization signal; bold: linker (1 NLS), (SEQ ID NO: 92)

[00123] *Cas9:human ADAT1 fusion (N-terminal)*

MDSLLMNRRKFLYQFKNVRWAKGRRETYLC*SMGTGTKCIGQSKMRKNGDILNDSHAEVIARR*
SFQRYLLHQLQLAATLKEDSIFVPGTQKGVWKLRRDLIFVFFSSHTPCGDASIIPMLEFEDQ
PCCPVFRNWAHNSSVEASSNLEAPGNERKCEDPDSPVTKKMRLEPGTAAREVTNGAAHHQSF
GKQKSGPISPGIHSCDLTVEGLATVTRIAPGSAKVIDVYRTGAKCVPGEAGDSGKPGAAFHQ
VGLLRVKPGRGDRTRSMSCSDKMARWNVLGCQGALLMHLLEEPIYLSAVVIGKCPYSQEAMQ
RALIGRCQNVSALPKGFGVQELKILQSDLLFEQSRSAVQAKRADSPGRLVPCGAAISWSAVP
EQPLDVTANGFPQGTTKKTIGSLQARSQISKVELFRSFQKLLSRIARDKWPHSLRVQKLDTY
*QEYKEAASSYQEAWSTLRKQVFGSWIRNPPDYHQF***GGGGSGGGGSGGGGS**DKKYSIGLAIGT
NSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR
KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH
LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN
PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED
AKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIK
RYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG
TEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFR
IPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLP

KHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKK
IECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIE
ERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQ
LIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPE
NIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNG
RDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKN
YWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKY
DENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE
SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNG
ETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVK
KDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNE
QKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTL
TNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID
NO: 35)
(underline: nuclear localization signal; bold: linker sequence)

[00124] *Cas9:human ADAT1 fusion (-terminal)*

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFK
VLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDD
SFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLA
LAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKS
RRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQ
IGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL
PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTR
KSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY
VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL
GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKR
RRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG
QGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
AIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT
KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLV
SDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKV
EKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK
RMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQI
SEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD**GGGGSGGGGS***SMGTGTKCIGQSKMRKNGD*
ILNDSHAEVIARRSFQRYLLHQLQLAATLKEDSIFVPGTQKGVWKLRRDLIFVFFSSHTPCG
DASIIPMLEFEDQPCCPVFRNWAHNSSVEASSNLEAPGNERKCEDPDSPVTKKMRLEPGTAA

REVTNGAAHHQSFGKQKSGPISPGIHSCDLTVEGLATVTRIAPGSAKVIDVYRTGAKCVPGE AGDSGKPGAAFHQVGLLRVKPGRGDRTRSMSCSDKMARWNVLGCQGALLMHLLEEPIYLSAV VIGKCPYSQEAMQRALIGRCQNVSALPKGFGVQELKILQSDLLFEQSRSAVQAKRADSPGRL VPCGAAISWSAVPEQPLDVTANGFPQGTTKKTIGSLQARSQISKVELFRSFQKLLSRIARDK WPHSLRVQKLDTYQEYKEAASSYQEAWSTLRKQVFGSWIRNPPDYHQF (SEQ ID NO: 36)(underline: nuclear localization signal; bold: linker sequence)

Example 2: Correction of a PI3K point mutation by a Cas9 fusion protein

[00125] An A3140G point mutation in exon 20 of the PI3KCA gene, resulting in an H1047R amino acid substitution in the PI3K protein is corrected by contacting a nucleic acid encoding the mutant protein with a Cas9:AID (SEQ ID NO: 30) or a Cas9:APOBEC1 (SEQ ID NO: 92) fusion protein and an appropriately designed sgRNA targeting the fusion protein to the mutation site in the encoding PI3KCA gene. The A3140G point mutation is confirmed via genomic PCR of the respective exon 20 sequence, *e.g.*, generation of a PCR amplicon of nucleotides 3000-3250, and subsequent sequencing of the PCT amplicon.

[00126] Cells expressing a mutant PI3K protein comprising an A3140G point mutation in exon 20 are contacted with an expression construct encoding the Cas9:AID (SEQ ID NO: 30) or a Cas9:APOBEC1 (SEQ ID NO: 92) fusion protein and an appropriately designed sgRNA targeting the fusion protein to the mutation site in the antisense strand of the encoding PI3KCA gene. The sgRNA is of the sequence 5'-
aucggaauctauuuugacucguuuuagagcuagaaaua
gcaaguuaaaauaaaggcuaguccguuaucaacuugaaaaaguggcaccgagucggugcuuuuu 3' (SEQ ID NO: 81); 5'-
ucggaaucuauuuugacucgguuuuagagcuagaaauagcaaguuaaaauaaaggcuaguccguuaucaacuugaaaaa
guggcaccgagucggugcuuuuu-3' (SEQ ID NO: 82); 5'-
cuuagauaaaacugagcaaggguuuuagagcuagaa
auagcaaguuaaaauaaaggcuaguccguuaucaacuugaaaaaguggcaccgagucggugcuuuuu-3' (SEQ ID NO: 83); 5'-
aucuauuuugacucguucucguuuuagagcuagaaauagcaaguuaaaauaaaggcuaguccguuaucaacuug
aaaaaguggcaccgagucggugcuuuuu-3' (SEQ ID NO: 84); 5'-
uaaaacugagcaagaggcuuguuuuagagcua
gaaauagcaaguuaaaauaaaggcuaguccguuaucaacuugaaaaaguggcaccgagucggugcuuuuu-3' (SEQ ID NO: 85); 5'-
ugguggcuggacaacaaaaaguuuuagagcuagaaauagcaaguuaaaauaaaggcuaguccguuaucaa
cuugaaaaaguggcaccgagucggugcuuuuu-3' (SEQ ID NO: 86); 5'-

gcuggacaacaaaaauggauguuuua
gagcuagaaauagcaaguuaaaauaaaggcuaguccguuaucaacuugaaaaaguggcaccgagucggugcuuuuu-3' (SEQ ID NO: 87); or 5'-
guguuaauuugucguacguaguuuuagagcuagaaauagcaaguuaaaauaaaggcua
guccguuaucaacuugaaaaaguggcaccgagucggugcuuuuu (SEQ ID NO: 88).

[00127] The cytosine deaminase activity of the Cas9:AID or the Cas9:APOBEC1 fusion protein results in deamination of the cytosine that is base-paired with the mutant G3140 to uridine. After one round of replication, the wild type A3140 is restored. Genomic DNA of the treated cells is extracted and a PCR amplicon of nucleotides 3000-3250 is amplified with suitable PCR primers. The correction of the A3140G point mutation after treatment of the cells with the fusion protein is confirmed by sequencing the PCR amplicon.

Example 3: Correction of a presenilin 1 point mutation by a Cas9 fusion protein

[00128] An A->G point mutation in codon 143 of the presenilin1 (PSEN1) gene, resulting in an I143V amino acid substitution in the PSEN1 protein is corrected by contacting a nucleic acid encoding the mutant PSEN1 protein with a Cas9:AID (SEQ ID NO: 30) or a Cas9:APOBEC1 (SEQ ID NO: 92) fusion protein and an appropriately designed sgRNA targeting the fusion protein to the mutation site in the encoding PSEN1 gene. See, *e.g.*, Gallo et. al., *J. Alzheimer's disease.* **2011;** 25: 425-431 for a description of an exemplary PSEN1 I143V mutation associated with familial Alzheimer's Disease. The A->G point mutation is confirmed via genomic PCR of the respective PSEN1 sequence, *e.g.*, generation of a PCR amplicon of about 100-250 nucleotides around exon 143, and subsequent sequencing of the PCT amplicon.

[00129] Cells expressing the mutant PSEN1 protein are contacted with an expression construct encoding the Cas9:AID (SEQ ID NO: 30) or a Cas9:APOBEC1 (SEQ ID NO: 92) fusion protein and an appropriately designed sgRNA targeting the fusion protein to the mutation site in the antisense strand of the encoding PSEN1 gene. The cytosine deaminase activity of the Cas9:AID or the Cas9:APOBEC1 fusion protein results in deamination of the cytosine that is base-paired with the mutant G in codon 143 to uridine. After one round of replication, the wild type A is restored. Genomic DNA of the treated cells is extracted and a PCR amplicon of 100-250 nucleotides is amplified with suitable PCR primers. The correction of the A->G point mutation after treatment of the cells with the fusion protein is confirmed by sequencing the PCR amplicon.

Example 4: Correction of an α_1-antitrypsin point mutation by a Cas9 fusion protein

[00130] A T->C point mutation in codon 55 of the α_1-antitrypsin gene, resulting in an L55P amino acid substitution in the α_1-antitrypsin protein is corrected by contacting a nucleic acid encoding the mutant α_1-antitrypsin protein with a Cas9:ADAT1 fusion protein (SEQ ID NO: 35 or 36) and an appropriately designed sgRNA targeting the fusion protein to the mutation site in the encoding α_1-antitrypsin gene. See, *e.g.*, Poller *et al.*, *Genomics.* **1993;** 17: 740-743 for a more detailed description of an exemplary codon 55 T->C mutation associated with chronic obstructive pulmonary disease (COPD). The T->C point mutation is confirmed via genomic PCR of the respective α_1-antitrypsin sequence encoding codon 55, *e.g.*, generation of a PCR amplicon of about 100-250 nucleotides, and subsequent sequencing of the PCT amplicon.

[00131] Cells expressing the mutant α_1-antitrypsin protein are contacted with an expression construct encoding the Cas9:AID (SEQ ID NO: 30) or a Cas9:APOBEC1 (SEQ ID NO: 92) fusion protein and an appropriately designed sgRNA targeting the fusion protein to the mutated nucleotide in codon 55 on the sense strand in the encoding α_1-antitrypsin gene. The cytosine deaminase activity of the Cas9:ADAT1 fusion protein results in deamination of the mutant cytosine to uridine thus correcting the mutation. Genomic DNA of the treated cells is extracted and a PCR amplicon of 100-250 nucleotides is amplified with suitable PCR primers. The correction of the T->C point mutation in codon 55 of the α_1-antitrypsin gene after treatment of the cells with the fusion protein is confirmed by sequencing the PCR amplicon

Example 5: Correction of a von Willebrand factor point mutation by a Cas9 fusion protein

[00132] A T->C point mutation in codon 509 of the von Willebrand factor gene, resulting in a C509A amino acid substitution in the von Willebrand factor protein is corrected by contacting a nucleic acid encoding the mutant von Willebrand factor protein with a Cas9:ADAT1 fusion protein (SEQ ID NO: 35 or 36) and an appropriately designed sgRNA targeting the fusion protein to the mutation site in the sense strand of the encoding von Willebrand factor gene. See, *e.g.*, Lavergne *et al.*, *Br. J. Haematol.* **1992;** 82: 66-7, for a description of an exemplary von Willebrand factor C509A mutation associated with von Willebrand disease (vWD). The T->C point mutation is confirmed via genomic PCR of the respective von Willebrand factor genomic sequence, *e.g.*, generation of a PCR amplicon of about 100-250 nucleotides around exon 509, and subsequent sequencing of the PCT

amplicon.

[00133] Cells expressing the mutant von Willebrand factor protein are contacted with an expression construct encoding the Cas9:ADAT1 fusion protein (SEQ ID NO: 35 or 36) and an appropriately designed sgRNA targeting the fusion protein to the mutation site in the sense strand of the encoding von Willebrand factor gene. The cytosine deaminase activity of the Cas9:ADAT1 fusion protein results in deamination of the mutant cytosine in codon 509 to uridine, thus correcting the mutation. Genomic DNA of the treated cells is extracted and a PCR amplicon of 100-250 nucleotides is amplified with suitable PCR primers. The correction of the T->C point mutation in codon 509 of the von Willebrand factor gene after treatment of the cells with the fusion protein is confirmed by sequencing the PCR amplicon.

Example 6: Correction of a Caspase 9 point mutation by a Cas9 fusion protein - neuroblastoma

[00134] A T->C point mutation in codon 197 of the Caspase-9 gene, resulting in an L197P amino acid substitution in the Caspase-9 protein is corrected by contacting a nucleic acid encoding the mutant Caspase-9 protein with a Cas9:ADAT1 fusion protein (SEQ ID NO: 35 or 36) and an appropriately designed sgRNA targeting the fusion protein to the mutation site in the sense strand of the encoding Caspase-9 gene. See, *e.g.*, Lenk *et al.*, *PLoS Genetics.* **2011;** 7: e1002104, for a description of an exemplary Caspase-9 L197P mutation associated with neuroblastoma (NB). The T->C point mutation is confirmed via genomic PCR of the respective Caspase-9 genomic sequence, *e.g.*, generation of a PCR amplicon of about 100-250 nucleotides around exon 197, and subsequent sequencing of the PCT amplicon.

[00135] Cells expressing the mutant Caspase-9 protein are contacted with an expression construct encoding the Cas9:ADAT1 fusion protein (SEQ ID NO: 35 or 36) and an appropriately designed sgRNA targeting the fusion protein to the mutation site in the sense strand of the encoding Caspase-9 gene. The cytosine deaminase activity of the Cas9:ADAT1 fusion protein results in deamination of the mutant cytosine in codon 197 to uridine, thus correcting the mutation. Genomic DNA of the treated cells is extracted and a PCR amplicon of 100-250 nucleotides is amplified with suitable PCR primers. The correction of the T->C point mutation in codon 197 of the Caspase-9 gene after treatment of the cells with the fusion protein is confirmed by sequencing the PCR amplicon.

Example 7: Deaminase activity of two dCas9-APOBEC1 fusion proteins

[00136] Two dCas9-APOBEC1 fusion proteins with different linkers were generated:

[00137] rAPOBEC1_GGS_dCas9:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHV
EVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHAD
PRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCII
LGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKGGSMDKKYSIGLAIGTNSV
GWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNR
ICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPIN
ASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKL
QLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD
EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEE
LLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIEC
FDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIH
DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIV
IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDM
YVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWR
QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEN
DKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF
VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG
GFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDL
IIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ
LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNL
GAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO:
94); underline = rAPOBEC1; double underline = dCas9.

[00138] rAPOBEC1_$(GGS)_3$_dCas9:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHV
EVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHAD
PRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCII
LGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKGGSGGSGGSMDKKYSIGLA
IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRY
TRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPT
IYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLF
EENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDL
AEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK
MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKIL

TFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK
VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDY
FKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRE
MIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRH
KPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMN
TKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYP
KLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDW
DPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK
EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPE
DNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHL
FTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD
(SEQ ID NO: 95); underline = rAPOBEC1; double underline = dCas9.

[00139] Deaminase activity of both fusion proteins were examined. A deaminase assay was adapted from Nuc. Acids Res. 2014, 42, p. 1095; J. Biol. Chem. 2004, 279, p 53379; J. Virology 2014, 88, p. 3850; and J. Virology 2006, 80, p. 5992, the entire contents of each of which are incorporated by reference.

[00140] Expression constructs encoding the fusion proteins were inserted into a CMV backbone plasmid (Addgene plasmid 52970; *see* Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* **2014;** 32(6): 577-82). The fusion proteins were expressed using a TNT Quick Coupled Transcription/Translation System (Promega). After 90 min, 5 μL of lysate was incubated with 5'-labeled ssDNA substrate (Cy3-ATTATTATTATTCCGCGGATTTATT TATTTATTTATTTATTT, SEQ ID NO: 96) and UDG (Uracil DNA Glycosylase) at 37 ℃ for 3 hr. A 1M solution of NaOH (10 μL) was then added to cleave the DNA at the abasic site. See Figure 4. The DNA was resolved on a 10% TBE PAGE gel (Figure 5). A negative control, where pUC19 was incubated in the TNT system, and a positive control, where the DNA has been synthesized with a "U" in place of the target C, were also included. Figure 5 illustrates that both fusion proteins exhibit cytosine deaminase activity.

REFERENCES

1. Humbert O, Davis L, Maizels N. Targeted gene therapies: tools, applications, optimization. *Crit Rev Biochem Mol.***2012;** 47(3):264-81. PMID: 22530743.

2. Perez-Pinera P, Ousterout DG, Gersbach CA. Advances in targeted genome editing. *Curr Opin Chem Biol.* **2012;** 16(3-4):268-77. PMID: 22819644.
3. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet.* **2010;** 11(9):636-46. PMID: 20717154.
4. Joung JK, Sander JD. TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol.* **2013;** 14(1):49-55. PMID: 23169466.
5. Charpentier E, Doudna JA. Biotechnology: Rewriting a genome. *Nature.* **2013;** 495, (7439):50-1. PMID: 23467164.
6. Pan Y, Xia L, Li AS, Zhang X, Sirois P, Zhang J, Li K. Biological and biomedical applications of engineered nucleases. *Mol Biotechnol.* **2013;** 55(1):54-62. PMID: 23089945.
7. De Souza, N. Primer: genome editing with engineered nucleases. *Nat Methods.* **2012;** 9(1):27. PMID: 22312638.
8. Santiago Y, Chan E, Liu PQ, Orlando S, Zhang L, Urnov FD, Holmes MC, Guschin D, Waite A, Miller JC, Rebar EJ, Gregory PD, Klug A, Collingwood TN. Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proc Natl Acad Sci U S A.* **2008;** 105(15):5809-14. PMID: 18359850.
9. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet.* **1999;** 22(3):231-8. PMID: 10391209.
10. Jansen R, van Embden JD, Gaastra W, Schouls LM. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol.* **2002;** 43(6):1565-75. PMID: 11952905.
11. Mali P, Esvelt KM, Church GM. Cas9 as a versatile tool for engineering biology. *Nat Methods.* **2013;** 10(10):957-63. PMID: 24076990.
12. Jore MM, Lundgren M, van Duijin E, Bultema JB, Westra ER, Waghmare SP, Wiedenheft B, Pul U, Wurm R, Wagner R, Beijer MR, Barendregt A, Shou K, Snijders AP, Dickman MJ, Doudna JA, Boekema EJ, Heck AJ, van der Oost J, Brouns SJ. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol.* **2011;** 18(5):529-36. PMID: 21460843.
13. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. *Science.* **2010;** 327(5962):167-70. PMID: 20056882.
14. Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. *Nature.* **2012;** 482(7385):331-8. PMID: 22337052.
15. Gasiunas G, Siksnys V. RNA-dependent DNA endonuclease Cas9 of the CRISPR system: Holy Grail of genome editing? *Trends Microbiol.* **2013;** 21(11):562-7. PMID: 24095303.
16. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell.* **2013;** 152(5):1173-83. PMID: 23452860.
17. Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, Thakore PI, Glass KA, Ousterout DG, Leong KW, Guilak F, Crawford GE, Reddy TE, Gersbach

CA. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods.* **2013;** 10(10):973-6. PMID: 23892895.

18. Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, Yang L, Church GM. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol.* **2013;** 31(9):833-8. PMID: 23907171.
19. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS, Qi LS. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell.* **2013;** 154(2):442-51. PMID: 23849981.
20. Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc.* **2013;** 8(11):2180-96. PMID: 24136345.
21. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. RNA-guided human genome engineering via Cas9. *Science.* **2013;** 339(6121):823-6. PMID: 23287722.
22. Cole-Strauss A, Yoon K, Xiang Y, Byrne BC, Rice MC, Gryn J, Holloman WK, Kmiec EB. Correction of the mutation responsible for sickle cell anemia by an RNA-DNA oligonucleotide. *Science.* **1996;** 273(5280):1386-9. PMID: 8703073.
23. Tagalakis AD, Owen JS, Simons JP. Lack of RNA-DNA oligonucleotide (chimeraplast) mutagenic activity in mouse embryos. *Mol Reprod Dev.* **2005;** 71(2):140-4. PMID: 15791601.
24. Ray A, Langer M. Homologous recombination: ends as the means. *Trends Plant Sci.* **2002;** 7(10):435-40. PMID 12399177.
25. Britt AB, May GD. Re-engineering plant gene targeting. *Trends Plant Sci.* **2003;** 8(2):90-5. PMID: 12597876.
26. Vagner V, Ehrlich SD. Efficiency of homologous DNA recombination varies along the Bacillus subtilis chromosome. *J Bacteriol.* **1988;** 170(9):3978-82. PMID: 3137211.
27. Saleh-Gohari N, Helleday T. Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human cells. *Nucleic Acids Res.* **2004;** 32(12):3683-8. PMID: 15252152.
28. Lombardo A, Genovese P, Beausejour CM, Colleoni S, Lee YL, Kim KA, Ando D, Urnov FD, Galli C, Gregory PD, Holmes MC, Naldini L. Gene editing in human stem cells using zince finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol.* **2007;** 25(11):1298-306. PMID: 17965707.
29. Conticello SG. The AID/APOBEC family of nucleic acid mutators. *Genome Biol.* **2008;** 9(6):229. PMID: 18598372.
30. Reynaud CA, Aoufouchi S, Faili A, Weill JC. What role for AID: mutator, or assembler of the immunoglobulin mutasome? *Nat Immunol.* **2003;** 4(7):631-8.
31. Bhagwat AS. DNA-cytosine deaminases: from antibody maturation to antiviral defense. *DNA Repair (Amst).* **2004;** 3(1):85-9. PMID: 14697763.
32. Navaratnam N, Sarwar R. An overview of cytidine deaminases. *Int J Hematol.* **2006;** 83(3):195-200. PMID: 16720547.

33. Holden LG, Prochnow C, Chang YP, Bransteitter R, Chelico L, Sen U, Stevens RC, Goodman MF, Chen XS. Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature.* **2008;** 456(7218):121-4. PMID: 18849968.
34. Chelico L, Pham P, Petruska J, Goodman MF. Biochemical basis of immunological and retroviral responses to DNA-targeted cytosine deamination by activation-induced cytidine deaminase and APOBEC3G. *J Biol Chem.* **2009;** 284(41). 27761-5. PMID: 19684020.
35. Pham P, Bransteitter R, Goodman MF. Reward versus risk: DNA cytidine deaminases triggering immunity and disease. *Biochemistry.* **2005;** 44(8):2703-15. PMID 15723516.
36. Barbas CF, Kim DH. Cytidine deaminase fusions and related methods. *PCT Int Appl.* **2010;** WO 2010132092 A2 20101118.
37. Chen X, Zaro JL, Shen WC. Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev.* **2013;** 65(10):1357-69. PMID: 23026637.
38. Gerber AP, Keller W. RNA editing by base deamination: more enzymes, more targets, new mysteries. *Trends Biochem Sci.* **2001;** 26(6):376-84. PMID: 11406411.
39. Yuan L, Kurek I, English J, Keenan R. Laboratory-directed protein evolution. *Microbiol Mol Biol Rev.* **2005;** 69(3):373-92. PMID: 16148303.
40. Cobb RE, Sun N, Zhao H. Directed evolution as a powerful synthetic biology tool. *Methods.* **2013;** 60(1):81-90. PMID: 22465795.
41. Bershtein S, Tawfik DS. Advances in laboratory evolution of enzymes. *Curr Opin Chem Biol.* **2008;** 12(2):151-8. PMID: 18284924.
42. Hida K, Hanes J, Ostermeier M. Directed evolution for drug and nucleic acid delivery. *Adv Drug Deliv Rev.* **2007;** 59(15):1562-78. PMID: 17933418.
43. Esvelt KM, Carlson JC, Liu DR. A system for the continuous directed evolution of biomolecules. *Nature.* **2011;** 472(7344):499-503. PMID: 21478873.
44. Husimi Y. Selection and evolution of bacteriophages in cellstat. *Adv Biophys.* **1989;** 25:1-43. PMID: 2696338.
45. Riechmann L, Holliger P. The C-terminal domain of TolA is the coreceptor for filamentous phage infection of E. coli. *Cell.* **1997;** 90(2):351-60. PMID: 9244308.
46. Nelson FK, Friedman SM, Smith GP. Filamentous phage DNA cloning vectors: a noninfective mutant with a nonpolar deletion in gene III. *Virology.* **1981;** 108(2):338-50. PMID: 6258292.
47. Rakonjac J, Model P. Roles of pIII in filamentous phage assembly. *J Mol Biol.* **1998;** 282(1):25-41.
48. Smith GP. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science.* **1985;** 228(4705):1315-7. PMID: 4001944.
49. Sheridan C. Gene therapy finds its niche. *Nat Biotechnol.* **2011;** 29(2):121-8. PMID: 21301435.
50. Lee JW, Soung YH, Kim SY, Lee HW, Park WS, Nam SW, Kim SH, Lee JY, Yoo NJ, Lee SH. PIK3CA gene is frequently mutated in breast carcinomas and hepatocellular carcinomas. *Oncogene.* **2005;** 24(8):1477-80. PMID: 15608678.
51. Ikediobi ON, Davies H, Bignell G, Edkins S, Stevens C, O'Meara S, Santarius T, Avis T, Barthorpe S, Brackenbury L, Buck G, Butler A, Clements J, Cole J, Dicks E, Forbes S, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Hunter C, Jenkinson A, Jones D, Kosmidou V, Lugg R, Menzies A, Mironenko T, Parker A, Perry J, Raine K, Richardson

D, Shepherd R, Small A, Smith R, Solomon H, Stephens P, Teaque J, Tofts C, Varian J, Webb T, West S, Widaa S, Yates A, Reinhold W, Weinstein JN, Stratton MR, Futreal PA, Wooster R. Mutation analysis of 24 known cancer genes in the NCI-60 cell line set. *Mol Cancer Ther.* **2006;** 5(11):2606-12. PMID: 17088437.

[00141] 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/or References sections, are hereby 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

[00142] Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents of the embodiments described herein. The scope of the present disclosure is not intended to be limited to the above description, but rather is as set forth in the appended claims.

[00143] 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 two or more members of a group are considered satisfied if one, more than one, or all of the group members are present, unless indicated to the contrary or otherwise evident from the context. The disclosure of a group that includes "or" between two or more group members provides embodiments in which exactly one member of the group is present, embodiments in which more than one members of the group are present, and embodiments in which all of the group members are present. For purposes of brevity those embodiments have not been individually spelled out herein, but it will be understood that each of these embodiments is provided herein and may be specifically claimed or disclaimed.

[00144] It is to be understood that the invention encompasses all variations, combinations, and permutations in which one or more limitation, element, clause, or descriptive term, from one or more of the claims or from one or more relevant portion of the description, is introduced into another claim. For example, a claim that is dependent on another claim can be modified to include one or more of the 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 making or using the composition according to any of the methods of making or using disclosed herein or according to methods known in the art, if any, 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.

[00145] Where elements are presented as lists, *e.g.*, in Markush group format, it is to be understood that every possible subgroup of the elements is also disclosed, and that any element or subgroup of elements 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 an embodiment, product, or method is referred to as comprising particular elements, features, or steps, embodiments, products, or methods that consist, or consist essentially of, such elements, features, or steps, are provided as well. For purposes of brevity those embodiments have not been individually spelled out herein, but it will be understood that each of these embodiments is provided herein and may be specifically claimed or disclaimed.

[00146] 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 some embodiments, to the tenth of the unit of the lower limit of the range, unless the context clearly dictates otherwise. For purposes of brevity, the values in each range have not been individually spelled out herein, but it will be understood that each of these values is provided herein and may be specifically claimed or disclaimed. 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.

[00147] 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.

CLAIMS

What we claim is:

1. A fusion protein comprising

(i) a nuclease-inactive Cas9 domain; and

(ii) an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase domain.

2. The fusion protein of claim 1, wherein the apolipoprotein B mRNA-editing complex (APOBEC) family deaminase domain is an APOBEC1 family deaminase domain.

3. The fusion protein of claim 1, wherein the apolipoprotein B mRNA-editing complex (APOBEC) family deaminase domain is an activation-induced cytidine deaminase (AID).

4. The fusion protein of any one of claims 1-3, wherein the apolipoprotein B mRNA-editing complex (APOBEC) family deaminase -domain is fused to the N-terminus of the Cas9 domain.

5. The fusion protein of any one of claims 1-3, wherein the apolipoprotein B mRNA-editing complex (APOBEC) family deaminase domain is fused to the C-terminus of the Cas9 domain.

6. The fusion protein of any one of claims 1-4, wherein the Cas9 domain and the apolipoprotein B mRNA-editing complex (APOBEC) family deaminase domain are fused via a linker.

7. The fusion protein of claim 6, wherein the linker comprises a $(GGGGS)_n$ (SEQ ID NO: 91), a $(G)_n$, an $(EAAAK)_n$ (SEQ ID NO: 5), a $(GGS)_n$, an SGSETPGTSESATPES (SEQ ID NO: 93), or an $(XP)_n$ motif, or a combination of any of these, wherein n is independently an integer between 1 and 30.

8. A method of DNA editing, the method comprising contacting a DNA molecule with

(a) a fusion protein comprising a nuclease-inactive Cas9 domain and an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase domain; and

(b) an sgRNA targeting the fusion protein of (a) to a target DNA sequence of the DNA molecule;

wherein the DNA molecule is contacted with the fusion protein and the sgRNA in an amount effective and under conditions suitable for the deamination of a cytidine base of the DNA molecule.

9. The method of claim 8, wherein the apolipoprotein B mRNA-editing complex (APOBEC) family deaminase domain is an APOBEC1 family deaminase domain.

10. The method of claim 8, wherein the apolipoprotein B mRNA-editing complex (APOBEC) family deaminase domain is an activation-induced cytidine deaminase (AID).

11. The method of any one of claims 8-10, wherein the apolipoprotein B mRNA-editing complex (APOBEC) family deaminase domain is fused to the N-terminus of the Cas9 domain.

12. The method of any one of claims 8-10, wherein the apolipoprotein B mRNA-editing complex (APOBEC) family deaminase domain is fused to the C-terminus of the Cas9 domain.

13. The method of any one of claims 8-12, wherein the Cas9 domain and the apolipoprotein B mRNA-editing complex (APOBEC) family deaminase domain are fused via a linker.

14. The method of claim 13, wherein the linker comprises a $(GGGGS)_n$ (SEQ ID NO: 91), a $(G)_n$, an $(EAAAK)_n$ (SEQ ID NO: 5), a $(GGS)_n$, an SGSETPGTSESATPES (SEQ ID NO: 93), or an $(XP)_n$ motif, or a combination of any of these, wherein n is independently an integer between 1 and 30.

15. The method of any one of claims 8-14, 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.

16. The method of claim 15, wherein the target DNA sequence comprises a point mutation associated with a disease or disorder, and wherein the deamination corrects the point mutation.

17. The method of any one of claims 8-15, wherein the target 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 base results in a sequence that is not associated with a disease or disorder.

18. The method of any one of claims 15-17, 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.

19. The method of any one of claims 15-18, wherein the contacting is *in vivo* in a subject having or diagnosed with a disease or disorder.

20. The method of any one of claims 15-19, wherein the disease or disorder is cystic fibrosis, phenylketonuria, epidermolytic hyperkeratosis (EHK), Charcot-Marie-Toot disease type 4J, neuroblastoma (NB), von Willebrand disease (vWD), myotonia congenital, hereditary renal amyloidosis, dilated cardiomyopathy (DCM), hereditary lymphedema, familial Alzheimer's disease, Prion disease, chronic infantile neurologic cutaneous articular syndrome (CINCA), desmin-related myopathy (DRM), or a neoplastic disease associated with a mutant PI3KCA protein.

21. The method of any one of claims 8-20, wherein the target DNA sequence comprises a T→C and/or an A→G point mutation that results in an amino acid sequence mutation in the PI3KCA protein as compared to the wild type PI3K protein, and wherein the method results in the deamination of the mutant C base.

22. The method of claim 21, wherein the point mutation is an A3140G mutation resulting in an H1047R substitution.

23. The method of any one of claims 8-20, wherein the target DNA sequence comprises a T→C and/or an A→G point mutation that results in an amino acid sequence mutation in the PSEN1 protein as compared to the wild type PSEN1 protein, and wherein the method results in the deamination of the mutant C base.

24. The method of claim 23, wherein the PSEN1 protein comprises an I143V substitution caused by an A→G point mutation in codon 143 of the PSEN1 gene.

25. The method of claim 23 or 24, wherein the PSEN1 point mutation is associated with Alzheimer's disease.

26. The method of any one of claims 23-25, wherein the contacting results in deamination of the mutant cytidine residue in codon 143 of the PSEN1 gene, thus correcting the A→G point mutation.

27. The method of any one of claims 8-20, wherein the target DNA sequence comprises a T→C and/or an A→G point mutation that results in an amino acid sequence mutation in the α-antitrypsin protein as compared to the wild type α-antitrypsin protein, and wherein the method results in the deamination of the mutant C base.

28. The method of claim 27, wherein the α-antitrypsin protein comprises an L55P substitution caused by a T→C point mutation in codon 55 of the α-antitrypsin gene.

29. The method of claim 27 or 28, wherein the α-antitrypsin point mutation is associated with chronic obstructive pulmonary disease (COPD).

30. The method of any one of claims 27-29, wherein the contacting results in deamination of the mutant cytidine residue in codon 55 of the α-antitrypsin gene, thus correcting the T→C point mutation.

31. The method of any one of claims 8-20, wherein the target DNA sequence comprises a T→C and/or an A→G point mutation that results in an amino acid sequence mutation in the vWF protein as compared to the wild type vWF protein, and wherein the method results in the deamination of the mutant C base.

32. The method of claim 31, wherein the vWF protein comprises an C509A substitution caused by a T→C point mutation in codon 509 of the vWF gene.

33. The method of claim 32, wherein the vWF point mutation is associated with von Willebrand disease.

34. The method of claim 31 or 32, wherein the contacting results in deamination of the mutant cytidine residue in codon 509 of the vWF gene, thus correcting the T→C point mutation.

35. The method of any one of claims 8-20, wherein the target DNA sequence comprises a T→C and/or an A→G point mutation that results in an amino acid sequence mutation in the Caspase-9 protein as compared to the wild type Caspase-9 protein, and wherein the method results in the deamination of the mutant C base.

36. The method of claim 35, wherein the Caspase-9 protein comprises an L197P substitution caused by a T→C point mutation in codon 197 of the Caspase-9 gene.

37. The method of claim 35 or 36, wherein the Caspase-9 point mutation is associated with neuroblastoma.

38. The method of any one of claims 35-37, wherein the contacting results in deamination of the mutant cytidine residue in codon 197 of the Caspase-9 gene, thus correcting the T→C point mutation.

39. The method of any one of claims 8-38, wherein the method further comprises detecting the deamination of the nucleotide base, optionally wherein the detecting is via PCR.

40. The method of any one of claims 8-39, wherein the fusion protein is the fusion protein of any one of claims 1-7.

41. A kit comprising:

a nucleic acid construct, comprising

a sequence encoding a fusion protein of claim 1,

wherein the nuclease-inactive Cas9 domain and the apolipoprotein B mRNA-editing complex (APOBEC) family deaminase domain optionally, are joined by a linker; and

suitable reagents, buffers, and/or instructions for use in generating the fusion protein from said nucleic acid construct.

42. The kit of claim 41, wherein the apolipoprotein B mRNA-editing complex (APOBEC) family deaminase domain is fused to the N-terminus of the nuclease-inactive Cas9 domain.

43. The kit of claim 41, wherein the apolipoprotein B mRNA-editing complex (APOBEC) family deaminase domain is fused to the C-terminus of the nuclease-inactive Cas9 domain.

44. The kit of any one of claims 41-43, wherein the encoded linker comprises a $(GGGGS)_n$ (SEQ ID NO: 91), a $(G)_n$, an $(EAAAK)_n$ (SEQ ID NO: 5), a $(GGS)_n$, an SGSETPGTSESATPES (SEQ ID NO: 93), or an $(XP)_n$ motif, or a combination of any of these, wherein n is independently an integer between 1 and 30.

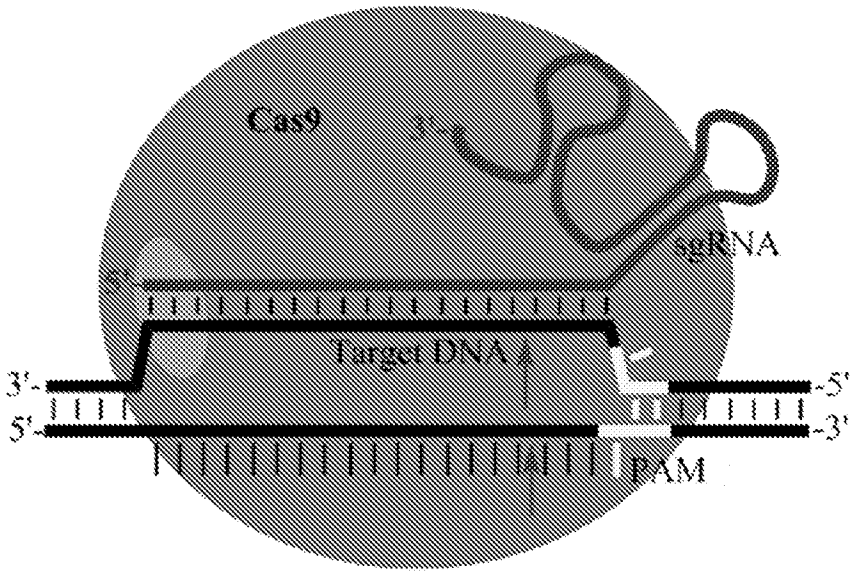


FIGURE 1

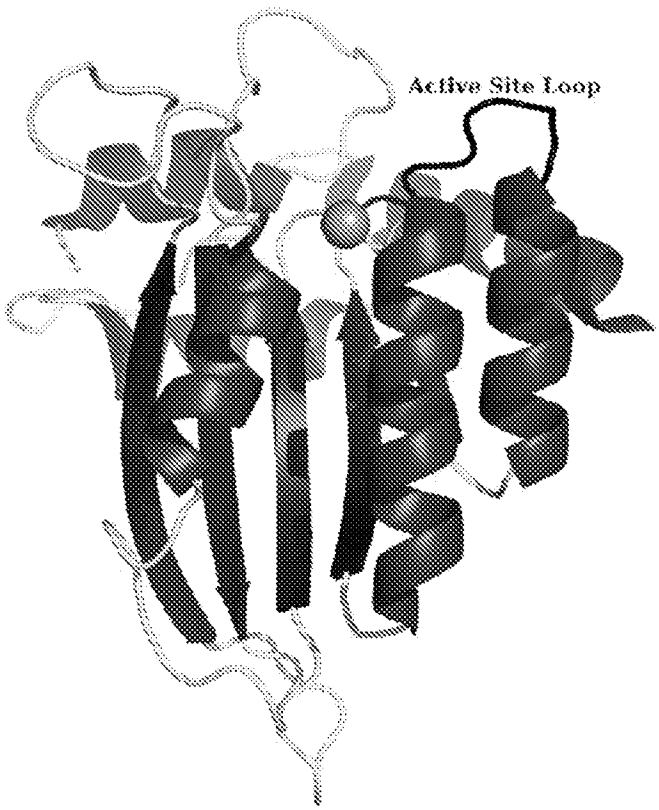


FIGURE 2

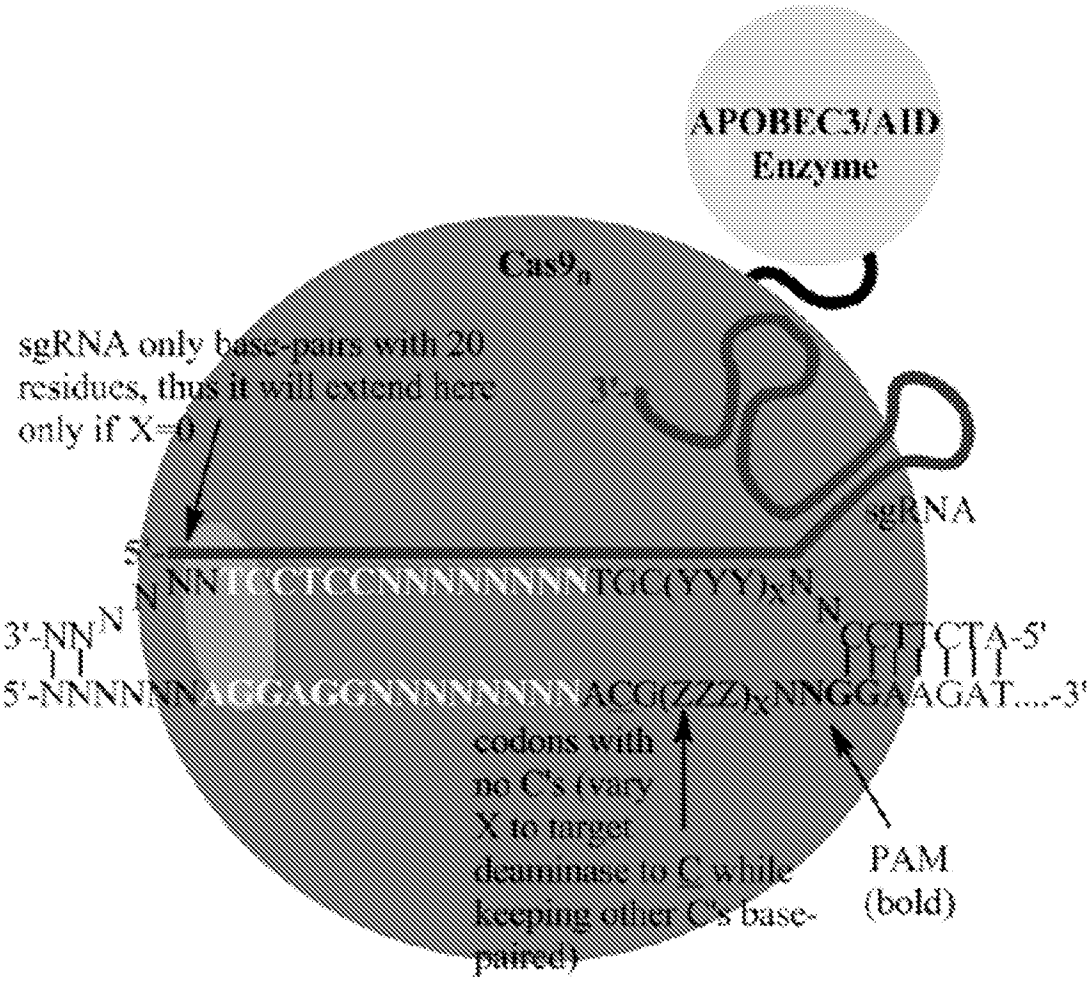

APOBEC3/AID
Enzyme
Cas9n
sgRNA only base-pairs with 20
residues, thus it will extend here
only if X=0
sgRNA
codons with
no C's (very
X to target
deaminase to C while
keeping other C's base-
paired)
PAM
(bold)

FIGURE 3

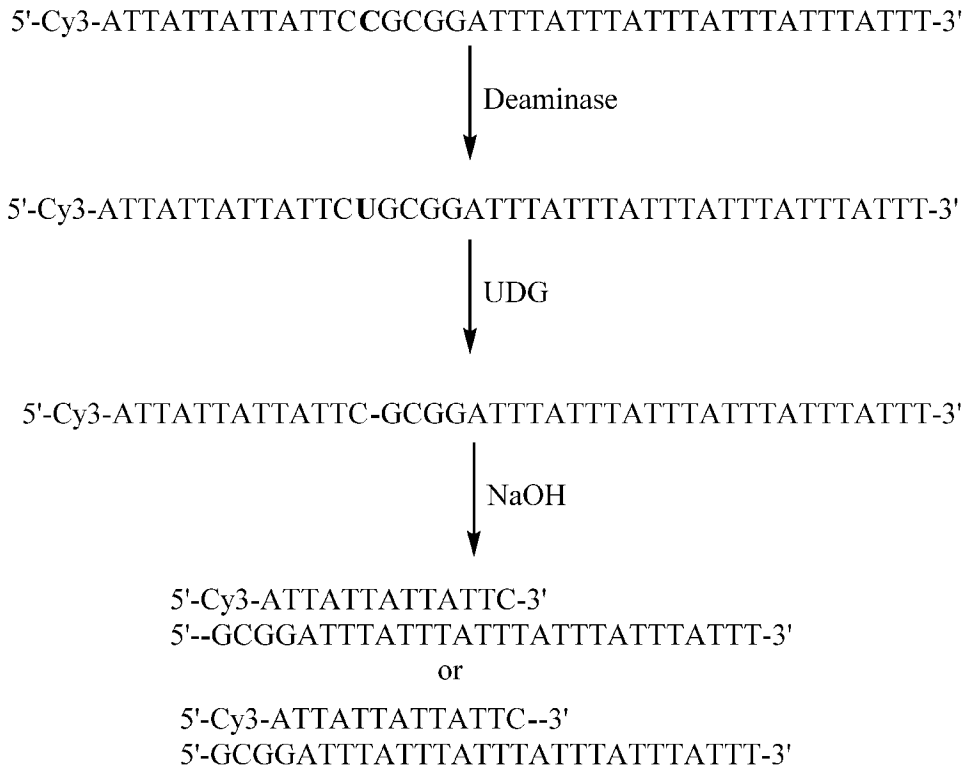

5'-Cy3-ATTATTATTATTCCGCGGATTTATTTATTTATTTATTTATTT-3'
Deaminase
5'-Cy3-ATTATTATTATTCUGCGGATTTATTTATTTATTTATTTATTT-3'
UDG
5'-Cy3-ATTATTATTATTC-GCGGATTTATTTATTTATTTATTTATTT-3'
NaOH
5'-Cy3-ATTATTATTATTC-3'
5'--GCGGATTTATTTATTTATTTATTTATTT-3'
or
5'-Cy3-ATTATTATTATTC--3'
5'-GCGGATTTATTTATTTATTTATTTATTT-3'

FIGURE 4

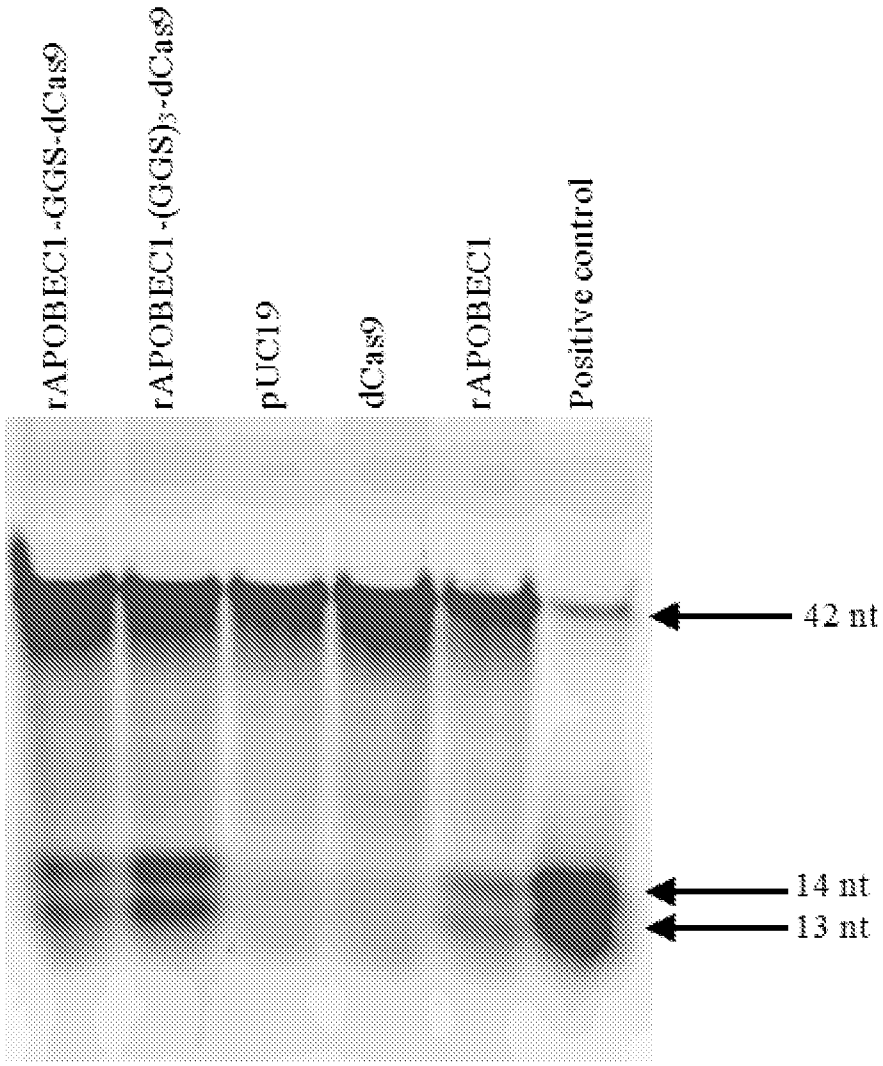

rAPOBEC1-GGS-dCas9
rAPOBEC1-(GGS)3-dCas9
pUC19
dCas9
rAPOBEC1
Positive control
42 nt
14 nt
13 nt

FIGURE 5