ENGINEERED TARGET SPECIFIC BASE EDITORS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The present application claims the benefit of U.S. Provisional Application No. 62/721,903, filed August 23, 2018; U.S. Provisional Application No. 62/753,696, filed October 31, 2018; U.S. Provisional Application No. 62/817,153, filed March 12, 2019; and U.S. Provisional Application No. 62/867,565, filed June 27, 2019, the disclosures of which are hereby incorporated by reference in their entireties.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on August 20, 2019, is named 8325-0180-S180-PCT_SL.txt and is 225,507 bytes in size.

TECHNICAL FIELD

[0003] The present disclosure is in the fields of polypeptide and genome engineering.

BACKGROUND

[0004] Artificial nucleases, such as engineered zinc finger nucleases (ZFN), transcription-activator like effector nucleases (TALENs), the CRISPR/Cas system with an engineered crRNA/tracr RNA ('single guide RNA'), also referred to as RNA guided nucleases, and/or nucleases based on the Argonaute system are revolutionizing the fields of medicine, biotechnology and agriculture. These molecular tools are allowing the genetic manipulation (*e.g.* editing) of genomes in organisms to a level never before possible. Artificial nucleases are capable of cleaving DNA such that following such cleavage, the cell is forced to 'heal' the break by either error-prone non-homologous end joining (NHEJ) or, in the presence of a substrate DNA with homology to the regions flanking the cut site, by insertion of the substrate DNA through homology-directed repair (HDR). Both of these processes start with a double strand break (DSB) in the DNA.

[0005] In some instances, engineered nucleases could possibly result in unwanted consequences (e.g. translocations, inversions and deletions) that may occur due to the induction of multiple DSB in the chromosome of a genetically-edited cell. For example, some evidence of chromosomal rearrangements including translocations, inversions and deletions have been observed following nuclease treatment (Kosicki, *et al.* (2018) *Nat Biotechnol* 36:765 and Shin, *et al.* (2017) *Nat Comm* doi:10.1038/ncomms15646), and more recently, there has been concern about induction of the p53 pathway following cleavage in some cells leading to apoptosis using the CRISPR/Cas system (Ihry, *et al.* (2018) *Nat Med* 24:939-946 and Haapaniemi, *et al.* (2018) *Nat Med* 24:927-930). Also, HDR typically is very inefficient in most eukaryotes, making gene correction difficult (Eid, *et al.* (2018) *Biochem J* 475:1955-1964).

[0006] In addition, Cas9 base editors such as AID-dCas9, APOBEC-dCas9 (e.g. APOBEC3G or APOBEC1), BE2, BE3 and BE4 (*see, e.g.,* Komor, *et al.* (2016) *Nature* 533:420-424; Komor, *et al.* (2017) *Science Advances* 3(8), eaao4774; Kim, *et al.* (2017) *Nat Biotechnol* 35(4):371-376) can exhibit a lack of specificity (see Kim, *et al.* (2019) *Nat Biotechnol* 10.1038/s41587-019-0050-1; Zuo, *et al.* (2019) *Science* DOI:10.1126/science.aav9973), rendering them unsuitable for a variety of purposes, including *in vivo* and *ex vivo* therapeutic applications.

[0007] Thus, there remains a need to accomplish genome (base) editing without inducing a double strand break and with high specificity.

SUMMARY

[0007a] In one aspect, provided herein is one or more expression vectors for expressing a system for editing an adenine or cytosine base in DNA, the system comprising: (i) an adenosine or cytidine deaminase, wherein the deaminase is a dimer comprising a first fusion protein comprising a first inactive domain of the deaminase and a first ZFP DNA-binding domain, and a second fusion protein comprising a second inactive domain of the deaminase and a second ZFP DNA-binding domain, wherein the first and second inactive domains dimerize to form an active deaminase; and (ii) a nickase fused to one or more ZFP DNA-binding domains.

[0007b] In another aspect, provided herein is one or more expression vectors for expressing a system for editing an adenine or cytosine base in DNA, the system comprising: (i) an adenosine or cytidine deaminase fused to one or more zinc finger

protein (ZFP) DNA-binding domains, and (ii) a zinc finger nuclease (ZFN) nickase, wherein the ZFN nickase is a dimer comprising a first polypeptide comprising a catalytically active nickase domain fused to a ZFP DNA-binding domain and a second polypeptide comprising a catalytically inactive nickase domain fused to another ZFP DNA-binding domain, wherein the first and second polypeptides dimerize to form an active ZFN nickase.

[0007c] In another aspect, provided herein is a kit for use in editing a base in DNA, the kit comprising the one or more expression vectors as described herein, optionally wherein the one or more expression vectors are one or more plasmid, viral, or mRNA vectors.

[0007d] In another aspect, provided herein is a cell comprising the one or more expression vectors as described herein, or a descendant of the cell wherein the cell comprises an edited base.

[0007e] In another aspect, provided herein is the one or more expression vectors as described herein for use in a method of editing a base in a target DNA in a cell, the method comprising providing to the cell said one or more expression vectors.

[0007f] In yet another aspect, provided herein is use of the cell as described herein or a descendant thereof in the manufacture of a medicament for treating a disease in a subject in need thereof, wherein said cell or descendant thereof is to be administered to the subject.

[0007g] In yet another aspect, provided herein is the cell as described herein or a descendant thereof for use in therapy.

[0008] The present disclosure provides methods and compositions to selectively edit DNA in a cell (for example, a base editor), including editing (*e.g.,* of a single base) without making a double-stranded cut in the target DNA (*e.g.,* the edited genome). Such base editors can be cytosine base editors (CBEs) which change a C:G to a T:A or adenine base editors (ABEs) which change A:T to G:C. Furthermore, because no double-stranded break is induced, there are no free DNA ends in the endogenous target and no translocations occur. Base editors as described herein can be used for gene knock out (*e.g.,* changing a regular codon into a stop codon, for instance using a cytosine base editor and/or mutating a splice acceptor site using either cytosine or adenine base editors); introducing mutations (*e.g.,* activating or repressing mutations) into a control element (*e.g.,* promoter region) of a gene;

correcting (reversing) disease-causing mutations (such as point mutations); and/or inducing mutations that that result in therapeutic benefits. The base editors as described herein may be provided (to a cell for *in vitro* or *ex vivo* uses or *in vivo* to a subject) for base editing in polypeptide and/or polynucleotide form. Among other advantages, the base editors of the invention can (1) increase specificity due to the additional DNA binding domain / length of the binding site an increased precision or targeting density due to reduced PAM requirements.; (2) expand (relax) PAM restrictions to allow targeting of sites not currently targetable; (3) increase editing efficiency at poorly performing PAM sites; and/or (4) improve efficiency at target sites targetable with non ZFP-anchored reagents and therefore supports a lower dose which then also results in lower off-target activity.

[0009] Thus, described herein are base editing compositions comprising at least one functional domain (*e.g.*, a DNA destabilizing molecule such as a nickase, a protein and/or a nucleotide) and at least one DNA-binding domain (*e.g.*, a zinc finger protein). In certain embodiments, the base editing composition edits an adenine (A) or cytidine (C) base in DNA, wherein the composition comprises: (1) at least one zinc finger protein (ZFP) DNA-binding domain; (2) at least one DNA destabilizing molecule; and (3) at least one adenine or cytosine deaminase, wherein the composition does not make a double-stranded cut in the DNA.

[0010] Any DNA destabilizing molecule may be used in the compositions described herein in any combination, including but not limited to a Cas9 nickase, a Cas9 protein (*e.g.*, dCas) operably linked to a single guide RNA (sgRNA), any RNA programmable system, a zinc finger nuclease nickase (ZFN nickase), a TALEN nickase, one or more proteins such as those shown in Table A, and/or one more nucleotides (*e.g.*, one or more peptide nucleic acids (PNAs), locked nucleic acids (LNAs) and/or bridged nucleic acids (BNAs)). In certain embodiments, the base editing composition comprises more than one DNA destabilizing molecule, for example one or more proteins (*e.g.*, Table A, nickases, etc.) and/or one or more nucleotides. In certain embodiments, the composition comprises a ZFN nickase and one or more additional proteins and/or nucleotide DNA destabilizing molecules (*e.g.*, one or more proteins of Table A and/or one or more nucleotides as described herein). In certain aspects, the base editing composition does not comprise a Cas9 protein, but may comprise other Cas protein (*e.g.*, non-Cas9 RNA programmable systems). In

certain embodiments, the DNA-destabilizing molecule comprises a zinc finger nuclease (ZFN) nickase.

[0011] The at least one zinc finger protein (ZFP) DNA-binding domain of the base editing composition may be operably linked to one or more of the other components of the base editing composition, for example to one or more of the DNA destabilizing molecules (*e.g.*, to Cas9 nickase, dCas9, etc.) and/or to the at least one adenine or cytosine deaminase. In certain embodiments, at least one ZFP DNA-binding domain is operably linked to the adenine or cytosine deaminase. In other embodiments, the base editing composition comprises first and second ZFP DNA-binding domains, wherein the first ZFP DNA-binding domain is operably linked to the Cas9 nickase. The ZFP DNA-binding domain may comprise 3, 4, 5, 6 or more fingers and may bind to a target site on either side (5' or 3') of the targeted base to be edited. In certain embodiments, the ZFP binds to a target site that is 1 to 100 (or any number therebetween) nucleotides on either side of the targeted base. In other embodiments, the ZFP binds to a target site that is 1 to 50 (or any number therebetween) nucleotides on either side of the targeted base.

[0012] Any adenine or cytosine deaminase can be used in the compositions described herein, including wild-type and/or evolved domains. In certain embodiments, the adenine or cytosine deaminase is comprised of first and second inactive domains that dimerize to form an active adenine or cytosine deaminase. In certain embodiments, the first inactive domain of the adenine or cytosine deaminase is operably linked to the Cas9 nickase and the second inactive domain of the adenine or cytosine deaminase is operably linked to a ZFP DNA-binding domain. In still further embodiments, the adenine or cytosine deaminase and the ZFP DNA-binding domain are both operably linked to the Cas9 nickase. In other embodiments, the base editor comprises first and second ZFP DNA-domains, the first ZFP operably linked to the Cas9 nickase and the second ZFP DNA-binding domain operably linked to the adenine or cytosine deaminase.

[0013] One or more polynucleotides encoding one or more base editing compositions as described herein are also provided. The polynucleotides may be carried on viral (*e.g.*, AAV, Ad, etc.) and/or non-viral (*e.g.*, plasmid, mRNA, etc.) vectors. Furthermore, a cell or population of cells comprising one or more compositions and/or the one or more polynucleotides as described herein are also

provided, as well as descendants of such cells, wherein the cells comprise an edited base.

[0014] Also provided are methods of editing a base in a target DNA (*e.g.*, DNA double stranded endogenous gene or extrachromosomal (episomal) sequence) using one or more of the compositions and/or polynucleotides as described herein. In certain embodiments, the methods comprise: (i) editing a cytidine base ("C") to a uracil base ("U"), optionally wherein the U is replaced with a thymidine base ("T") during DNA replication; (ii) editing an adenine base ("A") to an inosine ("I"), optionally wherein the I replaced with a guanine base ("G") during replication; and/or (iii) editing a CA or AC dinucleotide to a UI or an IU. In other embodiments, the editing in the cell results in: (i) changing a C:G base pair to an T:A base pair; (ii) changing a C:G base pair to a G:C base pair; (iii) changing an A:T base pair to a G:C base pair; (iv) introduction of a stop codon; and/or (v) editing or creating a splicing sequence. The methods may be used to correct any disease mutation (*e.g.*, point mutation), including in an exon or in an intron. wherein DNA in a chromosome or an extrachromosomal episome in the cell or the subject is edited. The method may be performed *in vitro, ex vivo,* or *in vivo.*

[0015] In one aspect, described herein are compositions and systems comprising a DNA-editing composition (*e.g.*, a base editing composition, also referred to herein as a base editing complex). The DNA-editing complex comprises at least one functional domain and a DNA-binding domain. In certain embodiments, the DNA-editing composition complex comprises a fusion molecule comprising a DNA-binding domain and, in addition, at least one DNA destabilizing molecule such as a nickase domain that makes a single-stranded cut in double-stranded DNA (*e.g.*, a DNA-nickase). In other embodiments, the DNA-editing composition (complex) comprises multiple (two or more) fusion molecules, for example a first catalytically active fusion molecule comprising a nickase including a first DNA-binding domain and nickase domain and a second catalytically inactive fusion molecule comprising a second DNA-binding domain and optionally one or more additional fusion molecules, each comprising an additional DNA-binding domain and one or more functional domains as described herein. In certain embodiments, the base editor comprises a composition as shown in any of Figures 1A through 1D. In certain embodiments, binding of the first and second (and optionally additional DNA binding domains) results in base-editing, for example when the catalytically active and catalytically

inactive fusion molecules dimerize. In some embodiments, the optional additional DNA binding domains bind to double stranded DNA, while in other embodiments, the DNA binding domains bind to single stranded DNA. In some embodiments, the DNA nickase is a ZFN nickase, a TALEN nickase or a CRISPR/Cas nickase, in which at least one functional (nickase) domain is operably linked to a DNA binding domain (*e.g.* a ZFP DNA binding domain, a TALE DNA binding domain and a sgRNA for use with a CRISPR/Cas system). In some embodiments, the DNA nickase (*e.g.*, fusion molecule) comprises a linker sequence between the nickase domain and the DNA binding domain. The nickase domain(s) may be positioned on either side of the DNA-binding domain, including at the N- or C-terminal side of the fusion molecule (N- and/or C-terminal to DNA-binding domain). In some embodiments, the linkers are selected from a bacterial selection system from a large linker library (>10e8 members). In some embodiments, the linkers range from four to 22 amino acid residues. In some embodiments, the linkers allow for specific positioning of a functional domain (for example a nickase domain) relative to a DNA binding domain (for example, linkage of the nickase domain to the N- or C- terminal side of the DNA binding domain). In some examples, the linker is selected using the method disclosed in Paschon, *et al.* (2019) *Nat Commun.* 10:1133. One or more polynucleotides (*e.g.*, constructs) encoding base editors (or components thereof) are also provided.

[0016] The DNA-editing complexes as described herein comprise one or more functional domains, including, but not limited to, one or more adenine deaminase domains, one or more cytidine deaminases, and/or one or more uracil DNA glycosylase inhibitors. One or more functional domains may be included in the catalytically active and/or the catalytically inactive fusion molecule of the DNA-editing complexes described herein. In some embodiments, the cytidine deaminase is an apolipoprotein B mRNA-editing complex 1 (APOBEC1) domain. In some embodiments, the cytidine deaminase is an Activation Induced Deaminase (AID). In some embodiments, the deaminase is an adenine deaminase. In some embodiments, the adenine deaminase is a wild-type or mutated (evolved) TadA (tRNA adenine deaminase (*see* Gaudelli, *et al.* (2017) *Nature* 551:464-471). In some embodiments, the adenine deaminase is ABE 7.8, ABE 7.9 or ABE 7.10 (Gaudelli, *ibid*) or ABEmax (Koblan, *et al.* (2018) *Nat Biotechnol.* 36(9):843-846). In some embodiments, the deaminase (adenine or cytidine) functional domain is assembled from two polypeptides comprising operably linked zinc fingers (*e.g.*, a split enzyme) or from

one or more ZFPs operably linked to one part of the split enzyme and a Cas9 nickase operably linked to the other component of the split enzyme (*see, e.g.,* Figure 1B). In some embodiments, assembly of the deaminase is driven by the binding of the operably linked zinc fingers to DNA targets such that the polypeptides are positioned to allow assembly. In some embodiments, the base editor further comprises a uracil DNA glycosylase inhibitor (UGI).

[0017] In one aspect, the base editor comprises a DNA-unwinding (also referred to as DNA-destabilizing) system derived from a CRISPR system, for instance a Cas9 (*e.g.,* naturally occurring and/or engineered Cas9) protein (*e.g.,* nickase) or a non-Cas9 protein. In certain embodiments, the base editor is a Cas9 base editor further comprises a zinc finger protein DNA-binding domain, which ZFP may be operably linked to any component of the Cas9 protein (*e.g.,* wild-type or engineered nickase) in any orientation, for example a base editor comprising a ZFP (a ZFP anchor) operably linked to the Cas9 protein, the sgRNA of the Cas9 nickase or the deamimase (wild-type or engineered (evolved) ABE or CBE). In certain embodiments, the ZFP is operably linked to the Cas9 domain of the base editor. In certain embodiments, the base editor comprises the components as shown in the Cas9 base editors of Figure 3.

[0018] In another aspect, the base editor does not comprise a DNA-unwinding (DNA-destabilizing) element derived from a Cas9 protein (also referred to as "Cas9-free"). In certain embodiments, the Cas9-free base editors of the invention comprise a ZFP-deaminase fusion protein and a ZFN nickase, and optionally one or more DNA-destabilizing factors. In certain embodiments, the DNA-destabilizing factor is a protein (*e.g.,* as shown in Table A) or an oligonucleotide (*e.g.,* one or more PNAs, LNAs and/or BNAs). The one or more non Cas9 DNA-destabilizing (unwinding) factor(s) (*e.g.,* proteins of Table A, LNAs, PNAs, BNAs, etc.) may be operably linked to any component of the base editor, for example either component of the ZFP-deaminase fusion protein and/or any of the components of the ZFN nickase. In some embodiments, the base editor comprises one or more protein and one or more nucleotide DNA-destabilizing (unwinding) factors. In still further embodiments, the Cas9-free base editors described herein comprise one or more proteins derived from a CRISPR system, which proteins are not Cas9 but have DNA-destabilizing (unwinding) properties.

[0019] In certain embodiments, the base editor comprises one or more nucleotide sequences, for example one or more DNA oligonucleotides, RNA oligonucleotides, peptide nucleic acids (PNAs), locked nucleic acids (LNAs) and/or bridged nucleic acids (BNAs), which can be used to provide a single stranded DNA substrate for base editors at the target site. This can be facilitated by e.g. duplex invasion, triplex invasion or a tail-clamp (Quijano, *et al.* (2017) *Yale J. Biol and Med.* 90:583-598; Pellestor and Paulasova (2004) *European J. Human Genetics* 12:694-700; Schleifman, *et al.* (2011) *Chem & Bio.* 18:1189-1198). The structure of the one or more nucleotide sequences of the base editor will vary in length; number and position of DNA and/or RNA and/or LNA and/or LNA and/or BNA bases; phosphorothioate bonds; other common modifications of these oligonucleotides depending on the target sequence composition.

[0020] In certain embodiments, the base editor comprises one or more PNAs, for example, gamma PNAs containing miniPEG substitutions and the gamma position for enhanced binding, increased solubility and improved delivery (Bahal, *et al.* (2014) *Current Gene Ther.* 14(5):331-342. In certain embodiments, the PNAs comprise one or more O indicates 8-amino-2,6-dioxaoctanoic acid linkers and/or one or more cytosines (C) or pseudoisocytosine residues. Optionally, one or more lysine (Lys) residues are included in the PNA, for example on the N- and/or C-terminals of the PNA sequence. In certain embodiments, 1, 2, 3, 4, 5 or more Lys residues are included at one or both terminals of the PNA. In certain embodiments, two or more PNAs are used in the base editor, for example in the same or reverse orientation relative to each other. In certain embodiments, the PNA comprises one or more PNAs as shown in Figures 8B to 8E, including but not limited to one or more PNAs of the structure: N-Lys-Lys-Lys-NNNNNNNNNN-OOO-NNNNNNNNNN-Lys-Lys-Lys-C; N-Lys-Lys-Lys-NNNNNNNNNN-OOO-NNNNNNNNNNNNNNN-Lys-Lys-Lys-C; N-Lys-Lys-Lys-NNNNNNNNNN-OOO-NNNNNNNNNN-Lys-Lys-Lys-C; N-Lys-Lys-Lys-NNNNNNNNNN-OOO-NNNNNNNNNN-Lys-Lys-Lys-C; and/or N-Lys-Lys-Lys-NNNNNNNNNNNNNNN-Lys-Lys-Lys-C, where O indicates 8-amino-2,6-dioxaoctanoic acid linkers and C indicates cytosine. The Lys resides on the N- and/or C-terminals of the PNA sequence are optional and pseudoisocytosine be can substituted for cytosine.

[0021] In other embodiments the base editor comprises one or more LNAs. LNAs can include a stacking linker and 2'-glycylamino-LNA for improved

performance (Geny, *et al.* (2016) *Nucleic Acids Res.* 44(5):2007-2019). In certain embodiments, the LNA comprise one or more phosphorothioate bonds, optionally between one or more LNA residues and/or DNA residues. In other embodiments, the LNA comprises one or more Cholesterol-TEG, which may increase uptake into cells. In certain embodiments, the base editor comprises one or more LNAs as shown in Figure 8F or 8G, including but not limited to one or more LNAs of the structure: 5`-NnNnNnNnNnNnNnNtctctnNnNnNnNnNnNnNnNnnNnnNnnNnnNn-3' (SEQ ID NO:1); 5'-N*n*NnNnNnNnNnNnNtctctnNnNnNnNnNnNnNnNnnNnnNnnNnn*N*n-3' (SEQ ID NO:69); and/or 5`-NnNnNnNnNnNnNnNtctctnNnNnNnNnNnNnNnNnnNnnNnnNnnNn-Chol-TEG-3' (SEQ ID NO:70), where LNA nucleotides are shown in uppercase; DNA nucleotides are in lower case; "*" indicates phosphorothioate bonds; and Chol-TEG indicates 3` Cholesterol-TEG for increased uptake into cells.

[0022] The components of the base-editing compositions described herein may be included in any combination (one or more nickase domains, one or more DNA-binding domains, one or more functional domains, etc.) and these components may be positioned in any order relative to each other. In some embodiments, the UGI, cytidine and/or adenine deaminase is(are) N-terminal of the DNA-binding domain of the catalytically inactive fusion molecule and/or N-terminal to the nickase domain of the catalytically active fusion molecule of the DNA-editing complex. In some embodiments, cytidine and/or adenine deaminase and/or UGI is(are) C-terminal of the DNA-binding domain of the catalytically inactive fusion molecule and/or C-terminal to the nickase domain of the catalytically active fusion molecule. In some embodiments, the one or more UGIs, cytidine and/or adenine deaminase(s) is(are) positioned between the DNA binding domain and the nickase domain(s) (in the catalytically active domain). In some embodiments, the fusion molecule comprises a cytidine deaminase and an adenine deaminase domain or a UGI, wherein the UGI, cytidine and adenine deaminases are positioned in any way with regard to the DNA-binding domain, each other and/or the nickase domain (*e.g.*, both N-terminal to the DNA-binding domain of the catalytically inactive fusion molecule in any order, both C-terminal to the DNA-binding domain of the catalytically inactive fusion molecule in any order, one N-terminal to the DNA-binding domain of the catalytically inactive fusion molecule, one C-terminal to the DNA-binding domain of the catalytically

inactive fusion molecule, N-terminal to the nickase domain and/or DNA-binding domain of the catalytically active fusion molecule, C-terminal to the nickase domain and/or DNA-binding domain of the catalytically active fusion molecule, one C-terminal to the nickase domain and/or DNA-binding domain of the catalytically active fusion molecule, one N-terminal to the nickase domain and/or DNA-binding domain of the catalytically active fusion molecule, between the nickase domain and the DNA-binding domain of the catalytically active fusion molecule, etc.). Non-limiting examples of configurations of one or more fusion molecules of the base-editing compositions are shown in the appended Figures and Examples. In some embodiments, the UGI, cytidine and/or adenine deaminase domains are linked to the other members of the DNA-editing complex using linkers known in the art. One or more polynucleotides encoding the base editors or components thereof are also provided.

[0023] In still further aspects, the DNA-editing complex comprises one or more functional domains comprising at least one uracil DNA glycosylase inhibitor (*e.g.* UGI) domain. The, which UGI domain(s) is(are) incorporated into the DNA-editing complex in any way such that the DNA-editing complex is operable. In some embodiments, the base editing complex comprises a bacteriophage Gam protein. In some embodiments, the base editing complex comprises a deaminase, a nickase, a UGI and/or a GAM protein. In some embodiments, the components of the base editing complex are provided in one, two or more gene expression constructs encoding one, two or more fusion proteins. In some embodiments, one or more uracil DNA glycosylase inhibitor domain(s) is/are linked to the other members of the complex using the linkers described above and known in the art. In some embodiments, a linker is used to link the uracil DNA glycosylase inhibitor to other members of the complex wherein the linker is identified using the method disclosed in Paschon, *et al.* (2019) *Nat Commun.* 10:1133.

[0024] In some embodiments, the DNA-editing (base editing) complex further comprises a molecule to assist in opening a double-strand DNA helix. In some embodiments, the molecule comprises an enzyme. In some embodiments, the enzyme is a helicase (for example, RecQ helicases (WRN, BLM, RecQL4 and RecQ5, (see Mo, *et al.* (2018) *Cancer Lett.* 413:1-10), DNA2 (Jia, *et al.* (2017) *DNA Repair* (Amst). 59:9-19) and any other eukaryotic helicases including for example, FANCJ, XPD, XPB, RTEL1, and PIF1 (Brosh (2013) *Nat Rev Canc* 13(8):542-558)). In some

embodiments, the enzyme is a bacterial and/or a viral helicase. Exemplary viral helicases include those encoded by the *Myoviridae* family of viruses (for example gp41, Dda, UvsW, Gene α, and Ban); those encoded by the *Podpviridae* family of viruses (for example 4B); those encoded by the *Siphoviridae, Baculoviridae, Herpesviridae, Polyomaviridae, Palillomaviridae* and *Poxviridae* families (for example, G40P, p143, UL5, UL9, Tag, E1, NPH-I, NPH-II, A18R, and VETF), or any other viral helicase known in the art (see *e.g.* Frick and Lam (2006) *Curr Pharm Des* 12(11):1315-1338). In some embodiments, the helicase enzyme is a bacterial enzyme. Exemplary bacterial helicases include the *P. aeruginosa* SF4 DnaB-like helicase, or the RecB and RecD helicases that are part of the bacterial RecBCD complex in bacteria such as *E coli* and *H. pylori* (Shadrick, *et al.* (2013) *J. Biomol Screen* 18(7):761-781). In some embodiments, engineered or evolved variants of multimeric helicases are used which result in monomeric helicase activity (see e.g. Brendza, *et al.* (2005) *PNAS* 102(29):10076-70081). In some embodiments, the molecule comprises a CRISPR/Cas complex. In some embodiments, the CRISPR/Cas complex comprises a guide RNA. In some embodiments, the complex comprises a Cas enzyme that is catalytically defective in its nuclease domains. In some embodiments, the complex comprises a Cas enzyme that is catalytically defective in one of its nuclease domains (for example a nickase). In some embodiments, the Cas enzyme is defective in its PAM recognition (Anders, *et al.* (2014) *Nature* 513(7519):569-573). In some embodiments, the Cas enzyme has relaxed (expanded) PAM requirements as compared to native PAM sequences (see for example Nishimasu, *et al.* (2018) *Science* 361:1259-1262). In certain embodiments, the Cas base editor as described herein exhibits relaxed (expanded) PAM requirements as compared to the NGG PAM sequence of SpCas9. In some embodiments, the molecule has helix-destabilizing properties. Exemplary helix-destabilizing molecules include ICP8 from herpes simplex virus type I (Boehmer and Lehman (1993) *J Virol* 67(2):711-715), Puralpha (Darbinian, *et al.* (2001) *J Cell Biochem* 80(4):589-95), and calf thymus DNA helix-destabilizing protein (Kohwi-Shigematsu, *et al.* (1978) *Proc Natl Acad Sci USA* 75(10):4689-93). In some embodiments, the molecule is involved in transcription and/or D loop formation/stabilization. Exemplary molecules of this class include Rad51, Rad52, RPA1, RPA2 and RPA3, Exo1, BLM, and HMGB1 and HMGB2. Other proteins that can be utilized include Bovin ROA1 and *E. coli* RecA or *E. coli* rad51. Other protein domains that may act as DNA helix destabilizers include the

RecI and Rec II domain from Cas9 or the RecII domain on its own, as well as any other helix destabilizing region from Cas9. Other non-limiting examples of suitable protein domains for use in the base editors described herein are shown in Table A.

[0025] In some embodiments, the molecule is a nucleic acid, including but not limited to oligonucleotides, PNAs, LNAs, BNAs and the like. In some embodiments, the nucleic acid is a DNA with homology to the region near the targeted editing. In some embodiments, the nucleic acid is an RNA with homology to the region near the targeted editing. In some embodiments, the RNA is modified. In some embodiments, the fusion molecule comprises amino acid linker sequences between one or more domains of the fusion molecule. In some embodiments, the molecule(s) used to assist in opening a double-strand DNA helix is/are linked to the other members of the DNA-editing complex using the linkers described above. In some embodiments, the molecule(s) used to assist in opening a double strand DNA helix is linked to the other members of the DNA editing complex is identified using known methods.

[0026] In certain embodiments, the nucleic acid comprises a PNA, for example a PNA comprising one or more O indicates 8-amino-2,6-dioxaoctanoic acid linkers and/or one or more cytosines (C) or pseudoisocytosine residues. Optionally, one or more lysine (Lys) residues are included in the PNA, for example on the N- and/or C-terminals of the PNA sequence. In certain embodiments, 1, 2, 3, 4, 5 or more Lys residues are included at one or both terminals of the PNA. In certain embodiments, two or more PNAs are used in the base editor, for example in the same or reverse orientation relative to each other. In certain embodiments, the one or more PNAs comprise: N-Lys-Lys-Lys-NNNNNNNNNN-OOO-NNNNNNNNNN-Lys-Lys-Lys-C; N-Lys-Lys-Lys-NNNNNNNNNN-OOO-NNNNNNNNNNNNNNN-Lys-Lys-Lys-C; N-Lys-Lys-Lys-NNNNNNNNNN-OOO-NNNNNNNNNN-Lys-Lys-Lys-C; N-Lys-Lys-Lys-NNNNNNNNNN-OOO-NNNNNNNNNN-Lys-Lys-Lys-C; and/or N-Lys-Lys-Lys-NNNNNNNNNNNNNNN-Lys-Lys-Lys-C (PNA #5), wherein O indicates 8-amino-2,6-dioxaoctanoic acid linkers and C indicates cytosine. The Lys resides on the N- and/or C-terminals of the PNA sequence are optional and pseudoisocytosine be can substituted for cytosine. *See, also,* Figures 8B to 8E.

[0027] In other embodiments the base editor comprises one or more LNAs. LNAs can include a stacking linker and 2'-glycylamino-LNA for improved performance (Geny, *et al.* (2016) *Nucleic Acids Res.* 44(5):2007-2019. In certain embodiments, the LNA comprise one or more phosphorothioate bonds, optionally

between one or more LNA residues and/or DNA residues. In other embodiments, the LNA comprises one or more Cholesterol-TEG, which may increase uptake into cells. In certain embodiments, the one or more LNAs comprise: 5`-NnNnNnNnNnNnNnNtctctnNnNnNnNnNnNnNnNnnNnnNnnNnnNn-3` (SEQ ID NO:1); 5`-N*n*NnNnNnNnNnNnNtctctnNnNnNnNnNnNnNnnNnnNnnNnn*N*n-3' (SEQ ID NO:69); and/or 5'-NnNnNnNnNnNnNnNtctctnNnNnNnNnNnNnNnnNnnNnnNnnNn-Chol-TEG-3' (SEQ ID NO:70), where LNA nucleotides are in uppercase; DNA nucleotides are in lower case; "*" indicates phosphorothioate bonds; and "Chol-TEG" indicates 3' Cholesterol-TEG for increased uptake into cells. *See, also,* Figures 8F and 8G.

[0028] These molecules may all be incorporated into the base editing system described herein, and may act to increase editing efficiency, decrease off target base editing, adjust the base editing window or alter the targeted type of nucleic acid base.

[0029] In some embodiments, functional domains as described herein are included in single fusion molecule. Alternatively, DNA-editing complexes that include multiple functional domains may be separated into separate fusion molecules in any way. In some embodiments, one fusion molecule comprises a DNA binding domain, a cytidine and/or adenine deaminase and a UGI, while a second fusion molecule comprises a nickase or half-nickase domain. In some embodiments, one fusion molecule comprises a catalytically inactive (dead) *Fok*I domain fused to a DNA binding domain fused to a deaminase domain, and the second fusion protein comprises a half *Fok*I nickase protein, a DNA binding domain and a UGI domain. In some embodiments, one fusion protein comprises a catalytically inactive (dead) *Fok*I domain fused to a deaminase domain fused to a UGI domain while a second fusion molecule comprises a functional nickase protein. In some embodiments, the one or more fusion proteins disclosed herein are fused in any order of domains within the fusion molecule that is operable. In some embodiments, the nickase domain is a Cas nickase domain, and in some embodiments, the nickase domain is a TALEN nickase domain. In some embodiments, one or more of the functional domains are linked to one or more other members of the complex using the linkers described above. In some embodiments, the one or more functional domains are linked to one or more other members of the complex using linkers identified using the methods disclosed in Paschon, *et al.* (2019) *Nat Commun.* 10:1133.

[0030] The base editor(s) described herein may be encoded by one or more polynucleotides. The one or more polynucleotides may be carried on viral vectors (AAV, Ad, etc.), non-viral vectors (plasmid, mRNA, etc.) or combinations thereof. In certain embodiments, one polynucleotide includes all the components of the base editor while in other embodiments, the components of the base editor are carried by two or more polynucleotides (*e.g.*, separate polynucleotides carrying split enzymes and/or ZFPs).

[0031] In another aspect, described herein are methods of editing (*e.g.*, gene editing) of a DNA molecule using one or more DNA-editing complexes as described herein. The methods described introducing one or more DNA-editing complexes into a cell such that the DNA molecule is edited. The cell may be isolated or may be in living subject (*e.g.*, via intravenous or other administration to the subject). In some embodiments, the DNA molecule is a chromosome or an extrachromosomal episome in a cell. In some embodiments, the chromosome or extrachromosomal episome comprises a cytidine base ("C") that is deaminated to a uracil base ("U") by the fusion protein disclosed herein. In some embodiments, the U is replaced with a thymidine base ("T") during DNA replication. In some embodiments, the chromosome or extrachromosomal episome comprises an adenine base ("A") that is deaminated to an inosine ("I") base by the fusion protein disclosed herein. In some embodiments, the I is replaced with a guanine base ("G") during replication. In some embodiments, the chromosome or extrachromosomal episome comprises an adenine and a cytidine base that are deaminated by the deaminases disclosed herein such that a CA or AC dinucleotide is deaminated into a UI or an IU dinucleotide (Figure 1 for exemplary systems).

[0032] In some embodiments, the nickase domain is derived from a *Fok*I DNA cleavage domain (see U.S. Patent Nos. 5,436,150; 8,703,489; 9,200,266; and 9,631,186). In some embodiments, the *Fok*I nickase comprises one or more mutations as compared to a parental *Fok*I nickase. Mutations as described herein, include but are not limited to, mutations that change the charge of the cleavage domain, for example mutations of positively charged residues to non-positively charged residues (*e.g.*, mutations of K and R residues (*e.g.*, mutated to S); N residues (*e.g.*, to D), and Q residues (*e.g.*, to E); mutations to residues that are predicted to be close to the DNA backbone based on molecular modeling and that show variation in *Fok*I homologs; and/or mutations at other residues (*e.g.*, U.S. Patent No. 8,623,618 and Guo, *et al.*

(2010) *J. Mol. Biol.* 400(1):96-107). Nickases can be ZFN nickases, TALEN nickases and CRISPR/Cas systems such as Cas nickases.

[0033] In some embodiments, the base editors comprise DNA-binding domains (*e.g.*, engineered nickase domains) comprising cleavage domains that are derived from *Fok*I or *Fok*I homologues and comprise a mutation in one or more of amino acid residues 416, 418, 422, 447, 448, 476, 479, 481 and/or 525, numbered relative to the wild-type full length *Fok*I as shown in SEQ ID NO:5, or corresponding residues in *Fok*I homologues. In some embodiments, the cleavage half domains derived from *Fok*I comprises a mutation in one or more of amino acid residues 414-426, 443-450, 467-488, 501-502, and/or 521-531, including one or more of 387, 393, 394, 398, 400, 416, 418, 422, 427, 434, 439, 441, 442, 444, 446, 448, 472, 473, 476, 478, 479, 480, 481, 487, 495, 497, 506, 516, 523, 525, 527, 529, 534, 559, 569, 570, and/or 571. The mutations may include mutations to residues found in natural restriction enzymes homologous to *Fok*I at the corresponding positions. In some embodiments, the mutations are substitutions, for example substitution of the wild-type residue with any different amino acid, for example alanine (A), cysteine (C), aspartic acid (D), glutamic acid (E), histidine (H), phenylalanine (F), glycine (G), asparagine (N), serine (S) or threonine (T). In some embodiments, the *Fok*I nuclease domain comprises a mutation at one or more of 416, 418, 422, 476, 447, 479, 481 and/or 525 (numbered relative to wild-type, SEQ ID NO:5). The nuclease domains may also comprise one or more mutations at positions 418, 432, 441, 448, 476, 481, 483, 486, 487, 490, 496, 499, 523, 527, 537, 538 and 559, including but not limited to ELD, KKR, ELE, KKS. *See, e.g.,* U.S. Patent No. 8,623,618. In some embodiments, the cleavage domain includes mutations at one or more of the residues 419, 420, 425, 446, 447, 470, 471, 472, 475, 478, 480, 492, 500, 502, 521, 523, 526, 530, 536, 540, 545, 573 and/or 574. In certain embodiments, the variant cleavage domains described herein include mutations to the residues involved in nuclease dimerization (dimerization domain mutations), and one or more additional mutations; for example to phosphate contact residues: *e.g.* dimerization mutants (such as ELD, KKR, ELE, KKS, etc.) in combination with one, two, three, four, five, six or more mutations at amino acid positions outside of the dimerization domain, for example in amino acid residues that may participate in phosphate contact. In some embodiments, the mutation at positions 416, 418, 422, 447, 448, 476, 479, 481 and/or 525 comprise replacement of a positively charged amino acid with an uncharged or a negatively

charged amino acid. In other embodiments, mutations at positions 446, 472 and/or 478 (and optionally additional residues for example in the dimerization or catalytic domains) are made. In some embodiments, the mutations comprise I479Q and/or Q481A mutations.

[0034] In some embodiments, the engineered cleavage half domain comprises mutations in the dimerization domain, for example, amino acid residues 490, 537, 538, 499, 496 and 486 in addition to the mutations described herein. In some embodiments, the invention provides fusion proteins wherein the engineered cleavage half-domain comprises a polypeptide in which the wild-type Gln (Q) residue at position 486 is replaced with a Glu (E) residue, the wild-type Ile (I) residue at position 499 is replaced with a Leu (L) residue and the wild-type Asn (N) residue at position 496 is replaced with an Asp (D) or a Glu (E) residue ("ELD" or "ELE") in addition to one or more mutations described herein. In some embodiments, the engineered nickase half domains are derived from a wild-type *Fok*I or *Fok*I homologue cleavage half domain and comprise mutations in the amino acid residues 490, 538 and 537, numbered relative to wild-type *Fok*I (SEQ ID NO:5) in addition to the one or more mutations at amino acid residues 416, 418, 422, 447, 448, 476, 479, 481 or 525. In some embodiments, the invention provides a fusion protein, wherein the engineered nickase half-domain comprises a polypeptide in which the wild-type Glu (E) residue at position 490 is replaced with a Lys (K) residue, the wild-type Ile (I) residue at position 538 is replaced with a Lys (K) residue, and the wild-type His (H) residue at position 537 is replaced with a Lys (K) residue or an Arg (R) residue ("KKK" or "KKR") (see U.S. Patent No. 8,962,281, incorporated by reference herein) in addition to one or more mutations described herein (*see* U.S. Patent Publication No. 2018/0087072).

[0035] In some embodiments, fusion molecules comprising a DNA binding domain and an engineered *Fok*I or homologue thereof cleavage half-domain as described herein that produce an artificial nuclease are provided. In some embodiments, the DNA-binding domain of the fusion molecule is a zinc finger binding domain (for example, an engineered zinc finger binding domain, ZFP). In some embodiments, the one or more of the zinc fingers are linked together using linkers identified using the methods disclosed in Paschon, *et al.*, *supra*. In some embodiments, the DNA-binding domain is a TALE DNA-binding domain (TALE). In some embodiments, the DNA binding domain comprises a DNA binding molecule

(*e.g.* guide RNA) and a catalytically inactive Cas or Cpf1 (also known as Cas12a) protein (for example dCas9 or dCpf1). In some embodiments, the DNA binding domain comprises a ZFP fused to a catalytically inactive Cas (dCas) protein. In some embodiments, the ZFP-dCas fusion protein comprises mutations to alter the PAM specificity. In some embodiments, the ZFP-dCas protein is not dependent on PAM recognition to specifically bind to a DNA sequence. In some embodiments, the DNA binding domain comprises a TALE fused to a dCas protein. In some embodiments, the TALE-dCas fusion protein comprises mutations to alter the PAM specificity. In some embodiments, the TALE-dCas protein is not dependent on PAM recognition to specifically bind to a DNA sequence. In any of the above embodiments, the linkers used to link the DNA binding domain (for example, ZFP, TALE or guide RNA and Cas system) to the engineered *Fok*I or homologue thereof are identified using the methods known in the art. *See, e.g.,* Paschon, *et al.* (2019) *Nat Commun.* 10:1133.

[0036] In some embodiments, the DNA-editing complex edits specific DNA bases in a double stranded DNA. In some embodiments, the edits are made in a DNA molecule within a cell. In some embodiments, the DNA is in a chromosome in a cell. In some embodiments, the editing results in the change from a C:G base pair to a T:A base pair. In some embodiments, the editing results in a change from a C:G base pair to a G:C base pair. In some embodiments, the editing results in a change from a A:T base pair to a G:C base pair. In some embodiments, the editing is done in an exon. In some embodiments, the editing results in the introduction of a stop codon (for example TAA, TAG, TGA). In some embodiments, the base editing results in the knock-out of gene expression of a targeted gene. In some embodiments, the editing is done in a sequence encoding a splicing sequence (for example, a U2 splice sequence wherein a 5' consensus sequence is G T A/G A/C/T G T/G/A/C A/G/T/C $(T/C/G/A)_5$ (SEQ ID NO:73) and the 3' consensus sequence is $(T/C)_{10}$ T/C/A/G C/T A G (SEQ ID NO:74); and a U12 splice sequence wherein a 5' consensus sequence is G/A T A T C T T/C and a 3' consensus sequence is $(T/G/A/G)_2$ T/A/C/G $(T/C/A/G)_2$ C/T A G/C, see Turunen, *et al.* (2013) *Wiley Interdiscip Rev RNA.* 4(1):61-76). In some embodiments, a new splicing sequence is created. In some embodiments, a splicing sequence is altered such that it no longer functions as a splicing sequence. In some embodiments, alteration of a splicing sequence causes exon skipping. In some embodiments, a sequence is altered such that a rare codon in created. In some embodiments, base editing causes correction of a point mutation in a DNA sequence

such that a gene associated with a disease is corrected. Non-limiting examples of base editing for treatment and/or prevention of disease include editing of JAK2 such that the V617F version is no longer expressed (thereby reducing activation of this gene which leads to uncontrolled blood cell production); base editing to knock out or repress other cancer genes such as BCR/ABL; base editing of A1AT; and the like. Exemplary diseases that may be treated include sickle cell disease, hemophilia, cystic fibrosis, phenylketonuria, Tay-Sachs, color blindness, Fabry disease, Friedreich's ataxia, prostate cancer, and many others.

[0037] In some embodiments, the base editing complexes as disclosed herein act on RNA molecules. In some embodiments, the base editors utilize an RNA-specific deaminase such as ADAR2 (adenosine deaminase acting on RNA type 2) (see Cox, *et al.* (2017) *Science* 358(6366):1019-1027).

[0038] Also disclosed herein are cells comprising any of the compositions (base-editing compositions and/or one or more polynucleotides encoding these compositions) as well as cells descended from these cells that have been modified by the methods and compositions disclosed herein. In some embodiments, the cell is a bacterial cell or a eukaryotic cell. In some embodiments, the cells comprise a base-editor complex and a base-editor complex induced DNA or RNA modification. The modified cells, and any cells derived from the modified cells do not necessarily comprise the base editor complex of the disclosure more than transiently, but the genomic modifications mediated by such base editor complexes remain.

[0039] In yet another aspect, methods for targeted editing of cellular chromatin in a region of interest; methods of treating infection; and/or methods of treating disease are disclosed herein. These methods maybe practiced *in vitro, ex vivo* or *in vivo* or a combination thereof. The methods involve editing cellular chromatin at a predetermined region of interest in cells by expressing a base editing complex as described herein (for example fusion polypeptides and optionally any associated nucleic acids in which one or more fusion polypeptide(s) comprise the engineered nickases as disclosed herein). In certain embodiments, the targeted editing of the on-target site is found in 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% or 100% of the cells.

[0040] The base editing complex as disclosed herein can be used in methods for targeted editing of cellular chromatin in a region of interest. Cells include cultured cells, cell lines, cells in an organism, cells that have been removed from an

organism for treatment in cases where the cells and/or their descendants will be returned to the organism after treatment, and cells removed from an organism, modified using the fusion molecules of the invention, and then returned to the organism in a method of treatment (cell therapy). A region of interest in cellular chromatin can be, for example, a genomic sequence or portion thereof.

[0041] A fusion molecule can be expressed in a cell, *e.g.*, by delivering the fusion molecule to the cell as a polypeptide, or by delivering a polynucleotide encoding the fusion molecule to a cell, wherein the polynucleotide, if DNA, is transcribed and is translated, to generate the fusion molecule. Further, if the polynucleotide is an mRNA encoding the fusion molecule, following delivery of the mRNA to the cell, the mRNA is translated, thus generating the fusion molecule.

[0042] In other aspects of the invention are provided methods and compositions for increasing base editing specificity. In some embodiments, methods are provided for increasing overall on-target editing specificity by decreasing off-target editing activity. In some embodiments, methods are provided for decreasing indel formation associated with base editing. In some embodiments, the engineered nickase components (nickase partners, for example a catalytically inactive ZFN partner and a catalytically active ZFN partner that form a ZFN nickase) of an engineered base editing complex are used to contact a cell, where each nickase partner of the complex is given in a ratio to the other partner other than one to one. In some embodiments, the ratio of the two partners is given at a 1:2, 1:3, 1:4, 1:5, 1:6, 1:8, 1:9, 1:10 or 1:20 ratio, or any value therebetween. In other embodiments, the ratio of the two partners is greater than 1:30. In some aspects, each partner is delivered to the cell as an mRNA or is delivered in a viral or non-viral vector where different quantities of mRNA or vector encoding each partner are delivered. In further embodiments, each partner of the nuclease complex may be comprised on a single viral or non-viral vector, but is deliberately expressed such that one partner is expressed at a higher or lower value that the other, ultimately delivering the cell a ratio of cleavage half domains that is other than one to one. In some embodiments, each cleavage half domain is expressed using different promoters with different expression efficiencies. In some embodiments, the two cleavage domains are delivered to the cell using a viral or non-viral vector where both are expressed from the same open reading frame, but the genes encoding the two partners are separated by a sequence (*e.g.* self-cleaving 2A sequence or IRES) that results in the 3' partner being expressed at a lower rate,

such that the ratios of the two partners are 1:2, 1:3, 1:4, 1:5, 1:6, 1:8, 1:9, 1:10 or 1:20 ratio, or any value therebetween. In other embodiments, the two partners are deployed at a ratio that is chosen to be different from 1:1.

[0043] In another aspect, described herein is a population of cells produced using one or more base editors as described herein. In certain embodiments, more than 5%-20% (or any value therebetween), preferably more than 20%, even more preferably more than 50% and even more preferably between 80% and 100% of the cells include the modification to the targeted base (*e.g.*, are base edited cells). In still further embodiments, the edited cells exhibit few or no off-target edits (unintended edits anywhere in the genome) and/or bystander (editing events in close proximity, for example 1-20 (or any value therebetween) nucleotides on either side of the intended target base, for example within the protospacer region of Cas9) mutations. Isolated populations of base edited cells as described herein can be used for *ex vivo* treatment of disease in a subject and/or can be further manipulated *ex vivo* (*e.g.*, via further rounds of base editing as described herein) prior to use as an *ex vivo* treatment. In addition, base editing can be conducted *in vivo* such that the disease or condition is treated in the subject following correction of the disease-related mutations *in vivo*.

[0044] In some embodiments, the nickase partners are fused to additional active domains. In some embodiments, the additional domains include one or more exemplary domains selected from one or more deaminases (for example A specific or C specific), a UGI domain, a helicase, and a GAM domain. In another aspect, described herein is a kit comprising a base editing complex as described herein or one or more polynucleotide(s) encoding one or more base editing complex proteins as described herein; ancillary reagents; and optionally instructions and suitable containers.

[0045] These and other aspects will be readily apparent to the skilled artisan in light of disclosure as a whole.

BRIEF DESCRIPTION OF THE DRAWINGS

[0046] **Figures 1A through 1D** are schematics depicting exemplary DNA-editing systems and complexes. Figure 1A shows systems comprising one catalytically inactive (indicated by "X") fusion molecule comprising a DNA-binding domain (*e.g.*, ZFP, TALE, sgRNA) and one catalytically active nickase fusion molecule (indicated by the scissors), also comprising a DNA-binding domain (*e.g.*,

ZFP, TALE, sgRNA). The catalytically active and inactive fusion molecules dimerize upon binding of the DNA-binding domains to their respective target sites and, following binding, edit the target DNA (*e.g.*, base editing). Figure 1A also shows complexes comprising two UGI domains. Figure 1B shows further exemplary Cas9 and Cas9-free systems for base editing. The top panel shows a base editor that functions via dimerization of the components of any adenosine or cytosine deaminase domains. The bottom panel of Figure 1B shows various embodiments of ABE and CBE base editors as described herein. Figure 1C shows another embodiment of a base editor as described herein comprising a Cas9 DNA destabilizing molecule (*e.g.*, RNA programmable system comprising dCas9 operably linked to a sgRNA), optionally linked to a ZFP anchor; a ZFP-deaminase fusion protein; and a ZFN nickase. In certain embodiments, the ZFN nickase is not present and the DNA destabilizing molecule comprises any RNA-programmable molecule. The schematic shows the ZFN nickase on the opposite side of the Cas9 nickase from the ZFP-deaminase fusion protein but it will be apparent the ZFN nickase and ZFP-deaminase can both be 3' or 5' to the Cas9 nickase. Figure 1D shows further Cas9-free (also referred to as non-Cas9) base editing systems. The triangle indicates where nicking occurs and "PNA" refers to peptide nucleic acid; "LNA" refers to locked nucleic acid and "BNA" refers to bridged nucleic acid. The nucleotides in these base editors (*e.g.*, DNA oligonucleotides, RNA oligonucleotides, peptide nucleic acids (PNAs), locked nucleic acids (LNAs) and/or bridged nucleic acids (BNAs)) can provide a single stranded DNA substrate for base editors at the target site.

[0047] **Figure 2** is a schematic showing DNA targeted by an exemplary adenine base editor. The drawing shows the DNA sequence near the A1AT Z mutation with a wildtype mRNA protospacer and PAM aligned on top (SEQ ID NO:78). To the right of the protospacer is shown the DNA targets of several different ZFPs. As shown, for ABE requiring a PAM sequence the target for base editing (also referred to as the base editing window) is typically 13-16 nucleotides from the PAM sequence and may be 3, 4, 5, 6, 7 or more nucleotides in size (shown in Figure is base editing window of 4 nucleotides) (SEQ ID NO:77).

[0048] **Figures 3A and 3B** are schematics depicting exemplary ZFP base editors. Figure 3A shows exemplary ZFP adenine base editors. The top panel shows an exemplary editor with the indicated components. The middle panel shows an exemplary ABE that uses two *E. coli* tRNA-specific adenosine deaminases (tadA),

where one is the wild type sequence and the other is an evolved sequence (Gaudelli, *et al.* (2017) *Nature* 551(7681):464-471)). The TadA domains are attached to each other and to the SPCas9 sequence using the linker shown (SEQ ID NO:2). The SpCas9 used is a VRVRFRR variant with a known relaxed PAM requirement (see Nishimasu, *et al.* (2018) *Science* 361:1259-1262). The Cas9 sequence is then linked to a ZFP DNA binding domain, where the linker used (SEQ ID NO:3) can comprise two NLS sequences and three HA tags. Cas9VR is also referred to as Cas9NG. Figure 3B shows exemplary Cas9 and Cas9-free base editors as described herein. The following abbreviations are used: "TadA" refers to wild-type adenine deaminase domain; TadA* refers to an evolved (engineered) adenine deaminase domain; "7.8" "7.9" "7.10" and "MAX" refer to evolved (engineered) adenine deaminase domains as described in Gaudelli, *et al.* (2017) *Nature* 551(7681):464-471) and Koblan, *et al.* (2018) *Nat Biotechnol.* 36(9):843-846; "SpCas9 [PAMs: NGG]" refers to Cas9 from Streptococcus pyogenes as described in Jinek, *et al.* (2012) *Science* 337(6096):816-21; "SpXCas9-3.7 [PAMs: NGN, GAA & GAT]" refers to a SpCas9 variant with broad PAM compatibility as described in Hu, *et al.* (2018) *Nature* 556(7699):57-63; "SpCas9-NG [PAMs: NGN; NAN in vitro]" refers to a SpCas9 variant with relaxed PAM requirements as described in Nishimasu, *et al.* (2018) *Science* 361(6408):1259-1262; "ScCas9 [PAMs: TGT, ...]" refers to a SpCas9 ortholog with minimal PAM specificity as described in Chatterjee, *et al.* (2018) *Sci Adv* 4(10):eaau0766. doi: 10.1126/sciadv.aau0766; "NO CAS9" means this domain is not present (Cas9-free base editor); "5F ZFP" refers to a five-finger ZFP; "6F ZFP" refers to a six finger ZFP; ">6F ZFP" refers to a ZFP having more than 6 fingers; "ZFP RQ" and "(...)" refers to modified ZFPs as described in Miller, *et al.* (2019) *Nature Biotechnology* 37(8):945-952

[0049] **Figure 4** is a schematic depicting adenine bases that lie within the editing window (SEQ ID NO:4) that are analyzed for targeting by the adenine base editor.

[0050] **Figures 5A through 5F** are schematics depict exemplary cytidine base editor construct. Figures 5A and 5B show base editor constructs comprising sequences encoding two UGI proteins linked to a ZFP DNA binding domain, further linked to sequences encoding either the APOBEC1 (Figure 5A) or AID (Figure 5B) cytidine enzymes capable of deaminating C nucleotides to U. Figures 5C and 5D depict two cytidine base editor constructs which lack the UGI domains of the

constructs of Figures 5A and 5B. Figures 5E and 5F depict two constructs that utilize sequences encoding a *Fok*I nickase. These constructs can be used as a pair where the sequences encoding the cytidine base editor are linked to a ZFP DNA binding domain which is then linked to a *Fok*I catalytically inactive nuclease domain. The second construct (Figure 5F) comprises sequences encoding two UGI domains linked to a ZFP DNA binding domain which is linked to sequences encoding a catalytically active *Fok*I nuclease domain. The pair may be constructed in any manner to make an active base editor wherein the active and inactive *Fok*I domains may be on either of the two partner constructs, and the UGI sequences and the cytidine base editor sequence can be on either partner.

[0051] **Figures 6A and 6B** depict two exemplary adenine base editors. Figure 6A shows a construct comprising sequences encoding two TadA domains, one wild type and one evolved, and linked to a ZFP DNA binding domain. As shown in Figure 6B, in some variations, the construct further comprises a catalytically inactive *Fok*I domain.

[0052] **Figures 7A through 7C** illustrate base editing of the JAK2 V617F target. Figure 7A shows the wildtype DNA double stranded sequence on the left (SEQ ID NO:30) with the encoded valine (V) indicated on top. The middle sequence shows the mutated DNA double strand sequence (SEQ ID NO:31) where the mutant phenylalanine (F) is indicated on the top. At the right is shown two possible base edited outcomes (SEQ ID NO:32 and 33) where the edited nucleotides are shown in bold with the changes to either a serine (S) or a proline (P) at the top. Figure 7B shows the DNA sequence (SEQ ID NO:34) surrounding the JAK2 V617F mutation, with the two closest PAM sites indicated. Figure 7B discloses the protein sequence as SEQ ID NO:79. Figure 7C shows exemplary results with the indicated base editors in K562 cells without the V617F mutation. Other A:T pairs within the base editing window were used to evaluate the activity of the tested base editors. ABEmax-Cas9NG indicates a Cas9NG nickase fused to ABEmax. ABEmax-Cas9 was anchored with 7 different ZFPs (shown are ZFP 2, ZFP 4, ZFP 6 and ZFP 7). Figure 7C shows the results for three different PAM sites (AAT, TAA, AAA; see Figure 7B) on the left. Here, both the ABEmax expression constructs as well as the corresponding sgRNAs were supplied as plasmid DNA (600 ng each). The ZFP anchored ABEmax-Cas9NG constructs show increased efficiency for all three PAM sites (approx. 2x for the AAT and AAA PAM sites; approx. 12x for the TAA PAM site). The base editors

for the AAT and TAA PAM sites were also tested at a higher dose (800 ng plasmid DNA) and show similar results. ZFP 6 results in approx. 2.5x higher activity for the AAT PAM site and ZFP 2 results in approximately 17x higher activity for the TAA PAM site.

[0053] **Figures 8A through 8G** are schematics showing exemplary base editors comprising nucleotides (*e.g.*, DNA oligonucleotides, RNA oligonucleotides, peptide nucleic acids (PNAs), locked nucleic acids (LNAs) and/or bridged nucleic acids (BNAs) used to provide a single stranded DNA substrate for base editors at the target site as shown in Figure 1D). Figure 8A depicts the targeted base to be edited ("X") before (left side) and after (right side) contact with the nucleotide-containing single stranded substrate of the base editor. Figure 8B depicts an exemplary PNA (PNA#1) for use in a base editor as described herein, the PNA having the structure: N-Lys-Lys-Lys-NNNNNNNNNN-OOO-NNNNNNNNNN-Lys-Lys-Lys-C. Figure 8C depicts an exemplary PNA (PNA#2) or use in a base editor as described herein, the PNA having the structure: N-Lys-Lys-Lys-NNNNNNNNNN-OOO-NNNNNNNNNNNNNNN-Lys-Lys-Lys-C. Figure 8D depicts an exemplary embodiment in which the base editor comprises 2 PNAs (PNA #3 having the structure N-Lys-Lys-Lys-NNNNNNNNNN-OOO-NNNNNNNNNN-Lys-Lys-Lys-C and PNA #4 having the structure N-Lys-Lys-Lys-NNNNNNNNNN-OOO-NNNNNNNNNN-Lys-Lys-Lys-C in reverse orientations relative to each other. Figure 8E depicts an exemplary embodiment in which the PNA comprises the structure N-Lys-Lys-Lys-NNNNNNNNNNNNNNN-Lys-Lys-Lys-C (PNA #5). In Figures 8B through 8E, O indicates 8-amino-2,6-dioxaoctanoic acid linkers and C indicates cytosine. The Lys resides on the N- and/or C-terminals of the PNA sequence are optional and pseudoisocytosine be can substituted for cytosine. Figures 8F and 8G depict exemplary embodiments of a base editor comprising an LNA. Figure 8F shows an exemplary LNA (LNA#1) (SEQ ID NO:80). Exemplary LNA#1 sequences include LNA#1a: 5'-NnNnNnNnNnNnNnNtctctnNnNnNnNnNnNnNnNnnNnnNnnNnnNn-3' (SEQ ID NO:1); LNA#1b: 5'-N*n*NnNnNnNnNnNnNtctctnNnNnNnNnNnNnNnNnnNnnNnnNnn*N*n-3' (SEQ ID NO:69); and LNA#1c: 5'-NnNnNnNnNnNnNnNtctctnNnNnNnNnNnNnNnNnnNnnNnnNnnNn-Chol-TEG-3' (SEQ ID NO:70). Figure 8G shows an exemplary embodiment in which the base

editor comprises 2 LNAs, shown in reverse orientation with respect to each other (LNA#2: 5'-NnNnNnNnNntctctNnNnNnNnNn-3' (SEQ ID NO:71) and LNA#3: 5'-NnNnNnNnNntctctNnNnNnNnNn-3' (SEQ ID NO:72)). In Figures 8F and 8G, LNA nucleotides are in uppercase; DNA nucleotides are in lower case; "*" indicates phosphorothioate bonds; and "Chol-TEG" indicates 3' Cholesterol-TEG for increased uptake into cells.

DETAILED DESCRIPTION

[0054] Artificial nucleases, such as engineered zinc finger nucleases (ZFN), transcription-activator like effector nucleases (TALENs), the CRISPR/Cas system with an engineered crRNA/tracr RNA ('single guide RNA'), also referred to as RNA guided nucleases, and/or nucleases based on the Argonaute system are revolutionizing the fields of medicine, biotechnology and agriculture. These molecular tools are allowing the genetic manipulation (e.g. editing) of genomes in organisms to a level never before possible. Artificial nucleases are capable of cleaving DNA such that following such cleavage, the cell is forced to 'heal' the break by either error-prone non-homologous end joining (NHEJ) or, in the presence of a substrate DNA with homology to the regions flanking the cut site, by insertion of the substrate DNA through homology-directed repair (HDR). Both of these processes start with a double strand break (DSB) in the DNA.

[0055] Described herein are compositions (systems) and methods for base editing that do not use a double-stranded cut for genetic modification. Base editing essentially relies on altering the identity of a specific base in a DNA strand and involved site-specific modification of the DNA base along with manipulation of the DNA repair machinery to avoid repair of the edited base. It is generally accomplished by using a system to open up the DNA double helix such that there are regions of single stranded DNA present. Next, the bases themselves are acted on by base modifying enzymes such as deaminases to change the nucleoside structure. For example, the Activation Induced Deaminase (AID) and apolipoprotein B mRNA editing enzyme catalytic polypeptide-like family proteins (APOBECs) are cytidine deaminases critical to antibody diversification and innate immunity against retroviruses. These enzymes convert cytidines (C) to uracils (U) in DNA. If DNA replication occurs before uracil repair, the replication machinery will treat the uracil

as thymine (T), leading to a C:G to T:A base pair conversion (Yang, *et al.* (2016) *Nat Commun* doi 10.1038/ncomms13330) so the system can be used to generate C to T point mutations.

[0056] Any of the base editors as described herein can be used for targeted base editing for any use, including but not limited to gene knockout (*e.g.*, alteration of a base to produce a stop codon in place of a regular codon; alteration of a base in a splice acceptor site); introduction of mutations in control (promoter) regions of a gene to activate or repress gene expression; and/or correction of disease-causing mutations by reversing a point mutation. Cells and cell lines comprising the base editors and/or targeted changes made by base editors (but no longer comprising the base editors themselves) are also provided.

[0057] The base editors of the present invention provide unexpectedly superior editing efficiencies and/or specificity as compared currently used based editors.

General

[0058] Practice of the methods, as well as preparation and use of the compositions disclosed herein employ, unless otherwise indicated, conventional techniques in molecular biology, biochemistry, chromatin structure and analysis, computational chemistry, cell culture, recombinant DNA and related fields as are within the skill of the art. These techniques are fully explained in the literature. *See*, for example, Sambrook *et al.* MOLECULAR CLONING: A LABORATORY MANUAL, Second edition, Cold Spring Harbor Laboratory Press, 1989 and Third edition, 2001; Ausubel *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, 1987 and periodic updates; the series METHODS IN ENZYMOLOGY, Academic Press, San Diego; Wolffe, CHROMATIN STRUCTURE AND FUNCTION, Third edition, Academic Press, San Diego, 1998; METHODS IN ENZYMOLOGY, Vol. 304, "Chromatin" (P.M. Wassarman and A. P. Wolffe, eds.), Academic Press, San Diego, 1999; and METHODS IN MOLECULAR BIOLOGY, Vol. 119, "Chromatin Protocols" (P.B. Becker, ed.) Humana Press, Totowa, 1999.

Definitions

[0059] The terms "nucleic acid," "polynucleotide," and "oligonucleotide" are used interchangeably and refer to a deoxyribonucleotide or ribonucleotide polymer, in linear or circular conformation, and in either single- or double-stranded form. For the purposes of the present disclosure, these terms are not to be construed as limiting with respect to the length of a polymer. The terms can encompass known analogues of natural nucleotides, as well as nucleotides that are modified in the base, sugar and/or phosphate moieties (*e.g.*, phosphorothioate backbones). In general, an analogue of a particular nucleotide has the same base-pairing specificity; *i.e.*, an analogue of A will base-pair with T.

[0060] The terms "polypeptide," "peptide," and "protein" are used interchangeably to refer to a polymer of amino acid residues. The term also applies to amino acid polymers in which one or more amino acids are chemical analogues or modified derivatives of a corresponding naturally-occurring amino acids.

[0061] "Binding" refers to a sequence-specific, non-covalent interaction between macromolecules (*e.g.*, between a protein and a nucleic acid). Not all components of a binding interaction need be sequence-specific (*e.g.*, contacts with phosphate residues in a DNA backbone), as long as the interaction as a whole is sequence-specific. Such interactions are generally characterized by a dissociation constant (K_d) of 10^{-6} M^{-1} or lower. "Affinity" refers to the strength of binding: increased binding affinity being correlated with a lower K_d. "Non-specific binding" refers to, non-covalent interactions that occur between any molecule of interest (*e.g.* an engineered nuclease) and a macromolecule (*e.g.* DNA) that are not dependent on-target sequence.

[0062] A "binding protein" is a protein that is able to bind non-covalently to another molecule. A binding protein can bind to, for example, a DNA molecule (a DNA-binding protein), an RNA molecule (an RNA-binding protein) and/or a protein molecule (a protein-binding protein). In the case of a protein-binding protein, it can bind to itself (to form homodimers, homotrimers, *etc.*) and/or it can bind to one or more molecules of a different protein or proteins. A binding protein can have more than one type of binding activity. For example, zinc finger proteins have DNA-binding, RNA-binding and protein-binding activity. In the case of an RNA-guided nuclease system, the RNA guide is heterologous to the nuclease component (Cas9 or Cfp1) and both may be engineered.

[0063] A "DNA binding molecule" is a molecule that can bind to DNA. Such DNA binding molecule can be a polypeptide, a domain of a protein, a domain within a

larger protein or a polynucleotide. In some embodiments, the polynucleotide is DNA, while in other embodiments, the polynucleotide is RNA. In some embodiments, the DNA binding molecule is a protein domain of a nuclease (*e.g.* the *Fok*I domain), while in other embodiments, the DNA binding molecule is a guide RNA component of an RNA-guided nuclease (*e.g.* Cas9 or Cpf1). DNA-binding molecules can comprise a protein, or a domain within a larger protein, that binds DNA in a sequence-specific manner, for example through one or more zinc fingers or through interaction with one or more ZFP recognition helix regions of a zinc finger protein (ZFP) or the RVDs of a TALE. DNA-binding molecules also include single guide RNA (sgRNA) of a CRISPR/Cas system and/or DNA-binding domains of a Ttago system. The term zinc finger DNA binding protein is often abbreviated as zinc finger protein or ZFP.

[0064] A "zinc finger DNA binding protein" (or binding domain) is a protein, or a domain within a larger protein, that binds DNA in a sequence-specific manner through one or more zinc fingers, which are regions of amino acid sequence within the binding domain whose structure is stabilized through coordination of a zinc ion. The term zinc finger DNA binding protein is often abbreviated as zinc finger protein or ZFP.

[0065] A "TALE DNA binding domain" or "TALE" is a polypeptide comprising one or more TALE repeat domains/units. The repeat domains are involved in binding of the TALE to its cognate target DNA sequence. A single "repeat unit" (also referred to as a "repeat") is typically 33-35 amino acids in length and exhibits at least some sequence homology with other TALE repeat sequences within a naturally occurring TALE protein. *See, e.g.*, U.S. Patent No. 8,586,526, incorporated by reference herein in its entirety.

[0066] DNA-binding domains can be "engineered" to bind to a predetermined nucleotide sequence, for example via engineering (altering one or more amino acids) of the recognition helix region of a naturally occurring zinc finger protein or by engineering of the amino acids involved in DNA binding (the "repeat variable diresidue" or RVD region). Therefore, engineered zinc finger proteins or TALE proteins are proteins that are non-naturally occurring. Non-limiting examples of methods for engineering zinc finger proteins and TALEs are design and selection. A designed protein is a protein not occurring in nature whose design/composition results principally from rational criteria. Rational criteria for design include application of substitution rules and computerized algorithms for processing information in a database storing information of existing ZFP or TALE designs and binding data. See,

for example, U.S. Patent Nos. 8,586,526; 6,140,081; 6,453,242; and 6,534,261; see also International Patent Publication Nos. WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536; and WO 03/016496.

[0067] A "selected" zinc finger protein, TALE protein or CRISPR/Cas system is not found in nature whose production results primarily from an empirical process such as phage display, interaction trap, rational design or hybrid selection. See *e.g.*, U.S. Patent Nos. 5,789,538; 5,925,523; 6,007,988; 6,013,453; and 6,200,759; and International Patent Publication Nos. WO 95/19431; WO 96/06166; WO 98/53057; WO 98/54311; WO 00/27878; WO 01/60970; WO 01/88197; and WO 02/099084.

[0068] "TtAgo" is a prokaryotic Argonaute protein thought to be involved in gene silencing. TtAgo is derived from the bacteria *Thermus thermophilus*. See, *e.g.* Swarts, *et al.* (2014) *Nature* 507(7491):258-261; Swarts, *et al.* (2012) *PLoS* One 7(4):e35888; G. Sheng, *et al.* (2013) *Proc. Natl. Acad. Sci.* U.S.A. 111, 652). A "TtAgo system" is all the components required including e.g. guide DNAs for cleavage by a TtAgo enzyme.

[0069] "Cleavage" refers to the breakage of the covalent backbone of a DNA molecule. Cleavage can be initiated by a variety of methods including, but not limited to, enzymatic or chemical hydrolysis of a phosphodiester bond. Both single-stranded cleavage and double-stranded cleavage are possible, and double-stranded cleavage can occur as a result of two distinct single-stranded cleavage events. DNA cleavage can result in the production of either blunt ends or staggered ends. In certain embodiments, fusion polypeptides are used for targeted double-stranded DNA cleavage.

[0070] A "cleavage half-domain" is a polypeptide sequence which, in conjunction with a second polypeptide (either identical or different) forms a complex having cleavage activity (preferably double-strand cleavage activity). The terms "first and second cleavage half-domains;" "+ and – cleavage half-domains" and "right and left cleavage half-domains" are used interchangeably to refer to pairs of cleavage half-domains that dimerize. The term "cleavage domain" is used interchangeably with the term "cleavage half-domain." The term "*Fok*I cleavage domain" includes the *Fok*I sequence as shown in SEQ ID NO:5 as well as any *Fok*I homologues.

[0071] An "engineered cleavage half-domain" is a cleavage half-domain that has been modified so as to form obligate heterodimers with another cleavage half-domain (*e.g.*, another engineered cleavage half-domain).

[0072] The term "editing" as used herein refers to a process wherein a nucleotide base is modified as compared to the initial (*e.g.*, wild-type) base at the same position. Base editing (*e.g.*, targeted point mutations) will necessarily reproduce the change in any mRNA that is transcribed from the edited DNA. Adenine and cytidine deaminases remove an amino group from their respective nucleotide targets, converting them into inosine and uridine respectively. During DNA repair or replication, inosine is recognized as guanine and uridine is recognized as thymine by polymerase enzymes, resulting in conversion of an A:T base pair into a G:C base pair, or C:G base pair into a T:A base pair in the double stranded DNA that has been edited. The "base editing window" refers to any bases that are subject to editing by the base editors as described herein may be any distance from any component of the editing system, typically within a region that is accessible following binding of at least one component of the base editing system to the target DNA. Base editors requiring a PAM sequence (*e.g.*, Cas9-containing editors) typically have a base editing window of 3, 4, 5, 6, 7 or more nucleotides that can be 13-16 or more nucleotides from the PAM sequence. Base editors as described herein can be used for targeted base editing for any use, including but not limited to gene knockout (*e.g.*, alteration of a base to produce a stop codon in place of a regular codon; alteration of a base in a splice acceptor site); introduction of mutations in control (promoter) regions of a gene to activate or repress gene expression; and/or correction of disease-causing mutations by reversing a point mutation. Cell lines comprising the base editors and/or targeted changes made by base editors (but no longer comprising the base editors themselves).

[0073] The term "sequence" refers to a nucleotide sequence of any length, which can be DNA or RNA; can be linear, circular or branched and can be either single-stranded or double stranded. The term "transgene" refers to a nucleotide sequence that is inserted into a genome. A transgene can be of any length, for example between 2 and 100,000,000 nucleotides in length (or any integer value therebetween or thereabove), preferably between about 100 and 100,000 nucleotides in length (or any integer therebetween), more preferably between about 2000 and 20,000 nucleotides in length (or any value therebetween) and even more preferable, between about 5 and 15 kb (or any value therebetween).

[0074] A "chromosome" is a chromatin complex comprising all or a portion of the genome of a cell. The genome of a cell is often characterized by its karyotype,

which is the collection of all the chromosomes that comprise the genome of the cell. The genome of a cell can comprise one or more chromosomes.

[0075] An "episome" is a replicating nucleic acid, nucleoprotein complex or other structure comprising a nucleic acid that is not part of the chromosomal karyotype of a cell. Examples of episomes include plasmids, minicircles and certain viral genomes. The liver specific constructs described herein may be episomally maintained or, alternatively, may be stably integrated into the cell.

[0076] An "exogenous" molecule is a molecule that is not normally present in a cell, but can be introduced into a cell by one or more genetic, biochemical or other methods. "Normal presence in the cell" is determined with respect to the particular developmental stage and environmental conditions of the cell. Thus, for example, a molecule that is present only during embryonic development of muscle is an exogenous molecule with respect to an adult muscle cell. Similarly, a molecule induced by heat shock is an exogenous molecule with respect to a non-heat-shocked cell. An exogenous molecule can comprise, for example, a functioning version of a malfunctioning endogenous molecule or a malfunctioning version of a normally-functioning endogenous molecule.

[0077] An exogenous molecule can be, among other things, a small molecule, such as is generated by a combinatorial chemistry process, or a macromolecule such as a protein, nucleic acid, carbohydrate, lipid, glycoprotein, lipoprotein, polysaccharide, any modified derivative of the above molecules, or any complex comprising one or more of the above molecules. Nucleic acids include DNA and RNA, can be single- or double-stranded; can be linear, branched or circular; and can be of any length. Nucleic acids include those capable of forming duplexes, as well as triplex-forming nucleic acids. See, for example, U.S. Patent Nos. 5,176,996 and 5,422,251. Proteins include, but are not limited to, DNA-binding proteins, transcription factors, chromatin remodeling factors, methylated DNA binding proteins, polymerases, methylases, demethylases, acetylases, deacetylases, kinases, phosphatases, ligases, deubiquitinases, integrases, recombinases, ligases, topoisomerases, gyrases and helicases.

[0078] An exogenous molecule can be the same type of molecule as an endogenous molecule, *e.g.*, an exogenous protein or nucleic acid. For example, an exogenous nucleic acid can comprise an infecting viral genome, a plasmid or episome introduced into a cell, or a chromosome that is not normally present in the cell.

Methods for the introduction of exogenous molecules into cells are known to those of skill in the art and include, but are not limited to, lipid-mediated transfer (*i.e.*, liposomes, including neutral and cationic lipids), electroporation, direct injection, cell fusion, particle bombardment, calcium phosphate co-precipitation, DEAE-dextran-mediated transfer and viral vector-mediated transfer. An exogenous molecule can also be the same type of molecule as an endogenous molecule but derived from a different species than the cell is derived from. For example, a human nucleic acid sequence may be introduced into a cell line originally derived from a mouse or hamster. Methods for the introduction of exogenous molecules into plant cells are known to those of skill in the art and include, but are not limited to, protoplast transformation, silicon carbide (*e.g.*, WHISKERS™), *Agrobacterium*-mediated transformation, lipid-mediated transfer (*i.e.*, liposomes, including neutral and cationic lipids), electroporation, direct injection, cell fusion, particle bombardment (*e.g.*, using a "gene gun"), calcium phosphate co-precipitation, DEAE-dextran-mediated transfer and viral vector-mediated transfer.

[0079] By contrast, an "endogenous" molecule is one that is normally present in a particular cell at a particular developmental stage under particular environmental conditions. For example, an endogenous nucleic acid can comprise a chromosome, the genome of a mitochondrion, chloroplast or other organelle, or a naturally-occurring episomal nucleic acid. Additional endogenous molecules can include proteins, for example, transcription factors and enzymes.

[0080] As used herein, the term "product of an exogenous nucleic acid" includes both polynucleotide and polypeptide products, for example, transcription products (polynucleotides such as RNA) and translation products (polypeptides).

[0081] A "fusion" molecule is a molecule in which two or more subunit molecules are linked, preferably covalently. The subunit molecules can be the same chemical type of molecule, or can be different chemical types of molecules. Examples of fusion molecules include, but are not limited to, fusion proteins (for example, a fusion between a protein DNA-binding domain and a cleavage domain), fusions between a polynucleotide DNA-binding domain (*e.g.*, sgRNA) operatively associated with a cleavage domain, and fusion nucleic acids (for example, a nucleic acid encoding the fusion protein).

[0082] Expression of a fusion protein in a cell can result from delivery of the fusion protein to the cell or by delivery of a polynucleotide encoding the fusion

protein to a cell, wherein the polynucleotide is transcribed, and the transcript is translated, to generate the fusion protein. Trans-splicing, polypeptide cleavage and polypeptide ligation can also be involved in expression of a protein in a cell. Methods for polynucleotide and polypeptide delivery to cells are presented elsewhere in this disclosure.

[0083] A "split enzyme" is an enzyme that has been split into two or more inactive polypeptide chains and then reassembled into an operable enzyme. The assembly of the split enzyme into an active protein often is driven by proximity where each inactive polypeptide chain is fused to other molecules that are capable of bringing the inactive chains physically together so that they can assemble, overcoming the entropic costs of fragmentation. The fused molecules can be other proteins that interact with each other, or any type of molecules that interact either with each other or with a common ligand, such that the interaction causes the assembly of the polypeptides that make up the split enzyme. *See* for example Shekhawat and Ghosh (2011) *Curr Opin Chem Biol* 15(6):789-797.

[0084] A "gene" for the purposes of the present disclosure, includes a DNA region encoding a gene product (see *infra*), as well as all DNA regions which regulate the production of the gene product, whether or not such regulatory sequences are adjacent to coding and/or transcribed sequences. Accordingly, a gene includes, but is not necessarily limited to, promoter sequences, terminators, translational regulatory sequences such as ribosome binding sites and internal ribosome entry sites, enhancers, silencers, insulators, boundary elements, replication origins, matrix attachment sites and locus control regions.

[0085] "Gene expression" refers to the conversion of the information contained in a gene, into a gene product. A gene product can be the direct transcriptional product of a gene (*e.g.*, mRNA, tRNA, rRNA, antisense RNA, ribozyme, structural RNA or any other type of RNA) or a protein produced by translation of an mRNA. Gene products also include RNAs which are modified, by processes such as capping, polyadenylation, methylation, and editing, and proteins modified by, for example, methylation, acetylation, phosphorylation, ubiquitination, ADP-ribosylation, myristilation, and glycosylation.

[0086] "Modulation" of gene expression refers to a change in the activity of a gene. Modulation of expression can include, but is not limited to, gene activation and gene repression. Genome editing (*e.g.*, cleavage, alteration, inactivation, random

mutation) can be used to modulate expression. Gene inactivation refers to any reduction in gene expression as compared to a cell that does not include a ZFP, TALE or CRISPR/Cas system as described herein. Thus, gene inactivation may be partial or complete.

[0087] A "region of interest" is any region of cellular chromatin, such as, for example, a gene or a non-coding sequence within or adjacent to a gene, in which it is desirable to bind an exogenous molecule. Binding can be for the purposes of targeted DNA cleavage and/or targeted recombination. A region of interest can be present in a chromosome, an episome, an organellar genome (*e.g.*, mitochondrial, chloroplast), or an infecting viral genome, for example. A region of interest can be within the coding region of a gene, within transcribed non-coding regions such as, for example, leader sequences, trailer sequences or introns, or within non-transcribed regions, either upstream or downstream of the coding region. A region of interest can be as small as a single nucleotide pair or up to 2,000 nucleotide pairs in length, or any integral value of nucleotide pairs.

[0088] "Eukaryotic" cells include, but are not limited to, fungal cells (such as yeast), plant cells, animal cells, mammalian cells and human cells (*e.g.*, T-cells), including stem cells (pluripotent and multipotent).

[0089] The terms "operative linkage" and "operatively linked" (or "operably linked") are used interchangeably with reference to a juxtaposition of two or more components (such as sequence elements), in which the components are arranged such that both components function normally and allow the possibility that at least one of the components can mediate a function that is exerted upon at least one of the other components. By way of illustration, a transcriptional regulatory sequence, such as a promoter, is operatively linked to a coding sequence if the transcriptional regulatory sequence controls the level of transcription of the coding sequence in response to the presence or absence of one or more transcriptional regulatory factors. A transcriptional regulatory sequence is generally operatively linked in *cis* with a coding sequence, but need not be directly adjacent to it. For example, an enhancer is a transcriptional regulatory sequence that is operatively linked to a coding sequence, even though they are not contiguous.

[0090] A "functional fragment" of a protein, polypeptide or nucleic acid is a protein, polypeptide or nucleic acid whose sequence is not identical to the full-length protein, polypeptide or nucleic acid, yet retains the same function as the full-length

protein, polypeptide or nucleic acid. A functional fragment can possess more, fewer, or the same number of residues as the corresponding native molecule, and/or can contain one or more amino acid or nucleotide substitutions. Methods for determining the function of a nucleic acid or protein (*e.g.*, coding function, ability to hybridize to another nucleic acid, enzymatic activity assays) are well-known in the art.

[0091] A polynucleotide "vector" or "construct" is capable of transferring gene sequences to target cells. Typically, "vector construct," "expression vector," "expression construct," "expression cassette," and "gene transfer vector" mean any nucleic acid construct capable of directing the expression of a gene of interest and which can transfer gene sequences to target cells. Thus, the term includes cloning, and expression vehicles, as well as integrating vectors.

[0092] The terms "subject" and "patient" are used interchangeably and refer to mammals such as human patients and non-human primates, as well as experimental animals such as rabbits, dogs, cats, rats, mice, and other animals. Accordingly, the term "subject" or "patient" as used herein means any mammalian patient or subject to which the expression cassettes of the invention can be administered. Subjects of the present invention include those with a disorder.

[0093] The terms "treating" and "treatment" as used herein refer to reduction in severity and/or frequency of symptoms, elimination of symptoms and/or underlying cause, prevention of the occurrence of symptoms and/or their underlying cause, and improvement or remediation of damage. Cancer, monogenic diseases and graft versus host disease are non-limiting examples of conditions that may be treated using the compositions and methods described herein.

[0094] "Chromatin" is the nucleoprotein structure comprising the cellular genome. Cellular chromatin comprises nucleic acid, primarily DNA, and protein, including histones and non-histone chromosomal proteins. The majority of eukaryotic cellular chromatin exists in the form of nucleosomes, wherein a nucleosome core comprises approximately 150 base pairs of DNA associated with an octamer comprising two each of histones H2A, H2B, H3 and H4; and linker DNA (of variable length depending on the organism) extends between nucleosome cores. A molecule of histone H1 is generally associated with the linker DNA. For the purposes of the present disclosure, the term "chromatin" is meant to encompass all types of cellular nucleoprotein, both prokaryotic and eukaryotic. Cellular chromatin includes both chromosomal and episomal chromatin.

[0095] An "accessible region" is a site in cellular chromatin in which a target site present in the nucleic acid can be bound by an exogenous molecule which recognizes the target site. Without wishing to be bound by any particular theory, it is believed that an accessible region is one that is not packaged into a nucleosomal structure. The distinct structure of an accessible region can often be detected by its sensitivity to chemical and enzymatic probes, for example, nucleases.

[0096] A "target site" or "target sequence" is a nucleic acid sequence that defines a portion of a nucleic acid to which a binding molecule will bind, provided sufficient conditions for binding exist. For example, the sequence 5'-GAATTC-3' is a target site for the Eco RI restriction endonuclease. An "intended" or "on-target" sequence is the sequence to which the binding molecule is intended to bind and an "unintended" or "off-target" sequence includes any sequence bound by the binding molecule that is not the intended target.

[0097] The terms "DNA destabilizing molecule" and "DNA unwinding molecule" are used interchangeably to refer to any molecule (*e.g.*, protein, nucleotide, small molecule, etc.) that aid in increasing the accessibility of (e.g., by exposing) the base targeted by the base editor. The term includes, but is not limited to, nickases, oligonucleotides (LNAs, PNAs, BNAs, etc.), RNA-programmable systems (*e.g.*, Cas proteins operably linked to sgRNAs), and other proteins (*e.g.*, Table A).

Base Editors

[0098] The base editing compositions (systems) described herein can directly change the identity of individual DNA base pairs without inducing double-stranded breaks. Thus, the base editors are not reliant on the DNA repair pathway preference to the target cell. Furthermore, because there is no double-stranded break made in the target DNA, there are no free DNA ends and, accordingly, no translocations.

[0099] The base editors described herein may be cytosine based editors (CBEs), which change a C:G pair to a T:A pair or adenine base editors (ABEs), which change an A:T pair to a G:C pair. These base editors can be used for inactivation (gene knock out) for example by turning regular codons into stop codons (*e.g.*, using a cytosine base editor) and/or by mutating splice acceptor sites using either cytosine or adenine base editors. In addition, base editors as described herein can be used for altering control (*e.g.*, promoter regions) of a gene to activate or repress expression of

the gene. Furthermore, base editing can be used to correct mutations, particularly disease-causing mutations.

[0100] Subsequent to the development of the first APOBEC-dCas9 base editors, a second base editor called BE2 was developed in which uracil DNA glycosylase (UGI) was added. Base excision repair is the cell's primary response to G:U mismatches and is initiated by excision of the uracil by uracil N-glycosylase (UNG). In an effort to protect the edited G:U intermediate from excision by UNG, a 83–amino acid uracil glycosylase inhibitor (UGI) was directly added to the C terminus of catalytically dead Cas9 (dCas9) resulting in an increase in efficiency (Komor, *et al.* (2017) *Science Advances* 3(8):eaao4774). In the early versions of base editors, a dead Cas9 was typically used such that the DNA replication machinery was used to carry out the final conversion of the nucleotide base opposite the edited base. In addition, Cas nickases have used to create a nick on the strand opposite of the one comprising the edited base. The creation of the nick attracts the DNA repair machinery such that the region downstream of the nick is excised and replaced using the edited strand as a template. The cytidine base editor BE3 used a Cas that was a nickase, Cas9 D10A, which also increased efficiency (Kim, *et al.* (2017) *Nat Biotechnol* 35(4):371-376). In yet another variant, the BE4 system uses two UGI domains, at both the N- and C-terminal ends of the complex for even greater efficiency. Another cytidine deaminase system relies on the activation-induced cytidine deaminase (AID) in combination with a nickase Cas9 ("target-AID").

[0101] When the base editor interacts with the DNA, the Cas-based editors require a PAM sequence to interact with, and then the window for activity (base editing window) is typically 13-16 bases from the 5' end of the PAM sequence (see, also, Figure 2). The activity window of the different editing systems described above vary. The target-AID system edits bases farther from the PAM sequence while the BE4 system edits those nearer to the PAM. Base editors have also been constructed based on the Cpf1 CRISPR system (Eid, *et al., ibid*). In addition, the BE4 configuration of base editors has been developed using both S. *pyogenes* and *S. aureus* derived CRISPR systems (BE4 and SaBE4 respectively). Therefore, Cas9 base editing systems are limited by the availability of a PAM sequence appropriately spaced from the target site (as the distance can significantly impact efficiency and/or specificity of the based editors) and/or the distance of the PAM sequence from the

base editing window (as shown in Figure 2A, 13-16 bases between the target base (*e.g.*, A) and the 5` end of the NGG PAM.

[0102] Thus, prior to the present invention, Cas9 base editors were known to induce genome-wide off-target effects as well as bystander effects (unintended edits near the targeted base). *See, e.g.*, Zuo, *et al.* (2019) *Science* 364(6437):289-292; Jin, *et al.* (2019) *Science* 364(6437):292-295; Gruenewald, *et al.* (2019) *Nature* 569(7756):433-437.

[0103] Additional base editor configurations have included the use of a GAM protein from bacteriophage Mu. In some instances, indels (insertions and deletions) have been observed as a result of some types of base editing. Because some base editors nick the strand opposite a U, cleavage of the glycosidic bond by UNG, followed by processing of the resulting apurinic or apyrimidinic site by AP lyase might result in a double stranded DNA break (DSB), potentially resulting in indel formation. The Gam protein of bacteriophage Mu binds to the ends of DSBs and protects them from degradation thus using Gam to bind the free ends of DSB may reduce indel formation during the process of base editing (Komor, *et al.* (2016) *Nature* 533:420-424; Komor, *et al.* (2017) *Science Advances* 3(8)). In addition to the cytidine deaminase editors, base editors have been developed with synthetic adenosine deaminases, which convert the adenine base into inosine (adenine base editors: "ABEs", see Gaudelli, *et al.* (2017) *Nature* 551(7681):464-471). Inosine can base pair with cytidine and subsequently corrected to guanine, thereby converting A into G, or A:T into G:C.

[0104] As described above, base editors as described herein can further comprise molecules that "open up" the DNA helix to expose the targeted base within a single stranded region of the DNA. Commonly known molecules that can accomplish this include but are not limited to DNA helicases, helix-destabilizing molecules and the bacterial DnaA protein, single-strand DNA binding proteins, triplex forming oligonucleotides or oligonucleotides.

[0105] In certain embodiments, the base editor comprises a protein domain that aids in unwinding (opening up) of the DNA helix to expose the targeted base for editing. Non-limiting examples of suitable proteins are shown in Table A.

Table A

Protein Name	Organism	Protein Sequence
ROA1	Bovine	MSKSESPKEPEQLRKLFIGGLSFETTDESLRSHFEQWGTLTDC VVMRDPNTKRSRGFGFVTYATVEEVDAAMNARPHKVDGRV VEPKRAVSREDSQRPGAHLTVKKIFVGGIKEDTEEHHLRDYFE QYGKIEVIEIMTDRGSGKKRGFAFVTFDDHDSVDKIVIQKYHT VNGHNCEVRKALSKQEMASASSSQRGRSGSGNFGGGRGGGF GGNDNFGRGGNFSGRGGFGGSRGGGGYGGSGDGYNGFGND GSNFGGGGSYNDFGNYNNQSSNFGPMKGGNFGGRSSGPYGG GGQYFAKPRNQGGYGGSSSSSSYGSGRRF (SEQ ID NO:35)
UvsW	BPT4	MDIKVHFHDFSHVRIDCEESTFHELRDFFSFEADGYRFNPRFR YGNWDGRIRLLDYNRLLPFGLVGQIKKFCDNFGYKAWIDPQI NEKEELSRKDFDEWLSKLEIYSGNKRIEPHWYQKDAVFEGLV NRRRILNLPTSAGKSLIQALLARYYLENYEGKILIIVPTTALTT QMADDFVDYRLFSHAMIKKIGGGASKDDKYKNDAPVVVGT WQTVVKQPKEWFSQFGMMMNDECHLATGKSISSIISGLNNC MFKFGLSGSLRDGKANIMQYVGMFGEIFKPVTTSKLMEDGQ VTELKINSIFLRYPDEFTTKLKGKTYQEEIKIITGLSKRNKWIA KLAIKLAQKDENAFVMFKHVSHGKAIFDLIKNEYDKVYYVSG EVDTETRNIMKTLAENGKGIIIVASYGVFSTGISVKNLHHVVL AHGVKSKIIVLQTIGRVLRKHGSKTIATVWDLIDSAGVKPKSA NTKKKYVHLNYLLKHGIDRIQRYADEKFNYVMKTVNLISFGP LEKKMLLEFKQFLYEASIDEFMGKIASCQTLEGLEELEAYYKK RVKETELKDTDDISVRDALAGKRAELEDSDDEVEESF (SEQ ID NO:36)
RECDL	*D. radiodurans*	MSAALPAEPFRVSGGVNKVRFRSDTGFTVMSATLRNEQGEDP DATVIGVMPPLDVGDTFSAEVLMEEHREYGYQYRVVNMVLE AMPADLSEEGVAAYFEARVGGVGKVLAGRIAKTFGAAAFDL LEDDPQKFLQVPGITESTLHKMVSSWSQQGLERRLLAGLQGL GLTINQAQRAVKHFGADALDRLEKDLFTLTEVEGIGFLTADK LWQARGGALDDPRRLTAAAVYALQLAGTQAGHSFLPRSRAE KGVVHYTRVTPGQARLAVETAVELGRLSEDDSPLFAAEAAA TGEGRIYLPHVLRAEKKLASLIRTLLATPPADGAGNDDWAVP KKARKGLSEEQASVLDQLAGHRLVVLTGGPGTGKSTTTKAV ADLAESLGLEVGLCAPTGKAARRLGEVTGRTASTVHRLLGY GPQGFRHNHLEPAPYDLLIVDEVSMMGDALMLSLLAAVPPG ARVLLVGDTDQLPPVDAGLPLLALAQAAPTIKLTQVVRQAAK NPIIQAAHGLLHGEAPAWGDKRLNLTEIEPDGGARRVALMVR ELGGPGAVQVLTPMRKGPLGMDHLNYHLQALFNPGEGGVRI AEGEARPGDTVVQTKNDYNNEIFNGTLGMVLKAEGARLTVD FDGNVVELTGAELFNLQLGYALTVHRAQGSEWGTVLGVLHE AHMPMLSRNLVYTALTRARDRFFSAGSASAWQIAAARQREA RNTALLERIRAH (SEQ ID NO:37)
RecA	*E. coli*	MAIDENKQKALAAALGQIEKQFGKGSIMRLGEDRSMDVETIS TGSLSLDIALGAGGLPMGRIVEIYGPESSGKTTLTLQVIAAAQR EGKTCAFIDAEHALDPIYARKLGVDIDNLLCSQPDTGEQALEI CDALARSGAVDVIVVDSVAALTPKAEIEGEIGDSHMGLAARM MSQAMRKLAGNLKQSNTLLIFINQIRMKIGVMFGNPETTTGG NALKFYASVRLDIRRIGAVKEGENVVGSETRVKVVKNKIAAP FKQAEFQILYGEGINFYGELVDLGVKEKLIEKAGAWYSYKGE KIGQGKANATAWLKDNPETAKEIEKKVRELLLSNPNSTPDFS VDDSEGVAETNEDF (SEQ ID NO:38)

RecB	*E. coli*	MSDVAETLDPLRLPLQGERLIEASAGTGKTFTIAALYLRLLLG LGGSAAFPRPLTVEELLVVTFTEAATAELRGRIRSNIHELRIAC LRETTDNPLYERLLEEIDDKAQAAQWLLLAERQMDEAAVFTI HGFCQRMLNLNAFESGMLFEQQLIEDESLLRYQACADFWRR HCYPLPREIAQVVFETWKGPQALLRDINRYLQGEAPVIKAPPP DDETLASRHAQIVARIDTVKQQWRDAVGELDALIESSGIDRR KFNRSNQAKWIDKISAWAEEETNSYQLPESLEKFSQRFLEDRT KAGGETPRHPLFEAIDQLLAEPLSIRDLVITRALAEIRETVARE KRRRGELGFDDMLSRLDSALRSESGEVLAAAIRTRFPVAMIDE FQDTDPQQYRIFRRIWHHQPETALLLIGDPKQAIYAFRGADIFT YMKARSEVHAHYTLDTNWRSAPGMVNSVNKLFSQTDDAFM FREIPFIPVKSAGKNQALRFVFKGETQPAMKMWLMEGESCGV GDYQSTMAQVCAAQIRDWLQAGQRGEALLMNGDDARPVRA SDISVLVRSRQEAAQVRDALTLLEIPSVYLSNRDSVFETLEAQE MLWLLQAVMTPERENTLRSALATSMMGLNALDIETLNNDEH AWDVVVEEFDGYRQIWRKRGVMPMLRALMSARNIAENLLA TAGGERRLTDILHISELLQEAGTQLESEHALVRWLSQHILEPDS NASSQQMRLESDKHLVQIVTIHKSKGLEYPLVWLPFITNFRVQ EQAFYHDRHSFEAVLDLNAAPESVDLAEAERLAEDLRLLYVA LTRSVWHCSLGVAPLVRRRGDKKGDTDVHQSALGRLLQKGE PQDAAGLRTCIEALCDDDIAWQTAQTGDNQPWQVNDVSTAE LNAKTLQRLPGDNWRVTSYSGLQQRGHGIAQDLMPRLDVDA AGVASVVEEPTLTPHQFPRGASPGTFLHSLFEDLDFTQPVDPN WVREKLELGGFESQWEPVLTEWITAVLQAPLNETGVSLSQLS ARNKQVEMEFYLPISEPLIASQLDTLIRQFDPLSAGCPPLEFMQ VRGMLKGFIDLVFRHEGRYYLLDYKSNWLGEDSSAYTQQAM AAAMQAHRYDLQYQLYTLALHRYLRHRIADYDYEHHFGGVI YLFLRGVDKEHPQQGIYTTRPNAGLIALMDEMFAGMTLEEA (SEQ ID NO:39)
RecC	*E. coli*	MLRVYHSNRLDVLEALMEFIVERERLDDPFEPEMILVQSTGM AQWLQMTLSQKFGIAANIDFPLPASFIWDMFVRVLPEIPKESA FNKQSMSWKLMTLLPQLLEREDFTLLRHYLTDDSDKRKLFQL SSKAADLFDQYLVYRPDWLAQWETGHLVEGLGEAQAWQAP LWKALVEYTHQLGQPRWHRANLYQRFIETLESATTCPPGLPS RVFICGISALPPVYLQALQALGKHIEIHLLFTNPCRYYWGDIKD PAYLAKLLTRQRRHSFEDRELPLFRDSENAGQLFNSDGEQDV GNPLLASWGKLGRDYIYLLSDLESSQELDAFVDVTPDNLLHNI QSDILELENRAVAGVNIEEFSRSDNKRPLDPLDSSITFHVCHSP QREVEVLHDRLLAMLEEDPTLTPRDIIVMVADIDSYSPFIQAV FGSAPADRYLPYAISDRRARQSHPVLEAFISLLSLPDSRFVSED VLALLDVPVLAARFDITEEGLRYLRQWVNESGIRWGIDDDNV RELELPATGQHTWRFGLTRMLLGYAMESAQGEWQSVLPYDE SSGLIAELVGHLASLLMQLNIWRRGLAQERPLEEWLPVCRDM LNAFFLPDAETEAAMTLIEQQWQAIIAEGLGAQYGDAVPLSL LRDELAQRLDQERISQRFLAGPVNICTLMPMRSIPFKVVCLLG MNDGVYPRQLAPLGFDLMSQKPKRGDRSRRDDDRYLFLEAL ISAQQKLYISYIGRSIQDNSERFPSVLVQELIDYIGQSHYLPGDE ALNCDESEARVKAHLTCLHTRMPFDPQNYQPGERQSYAREW LPAASQAGKAHSEFVQPLPFTLPETVPLETLQRFWAHPVRAFF QMRLQVNFRTEDSEIPDTEPFILEGLSRYQINQQLLNALVEQD DAERLFRRFRAAGDLPYGAFGEIFWETQCQEMQQLADRVIAC RQPGQSMEIDLACNGVQITGWLPQVQPDGLLRWRPSLLSVAQ GMQLWLEHLVYCASGGNGESRLFLRKDGEWRFPPLAAEQAL HYLSQLIEGYREGMSAPLLVLPESGGAWLKTCYDAQNDAML

		DDDSTLQKARTKFLQAYEGNMMVRGEGDDIWYQRLWRQLT PETMEAIVEQSQRFLLPLFRFNQS (SEQ ID NO:40)
RecD	*E. coli*	MKLQKQLLEAVEHKQLRPLDVQFALTVAGDEHPAVTLAAAL LSHDAGEGHVCLPLSRLENNEASHPLLATCVSEIGELQNWEE CLLASQAVSRGDEPTPMILCGDRLYLNRMWCNERTVARFFNE VNHAIEVDEALLAQTLDKLFPVSDEINWQKVAAAVALTRRIS VISGGPGTGKTTTVAKLLAALIQMADGERCRIRLAAPTGKAA ARLTESLGKALRQLPLTDEQKKRIPEDASTLHRLLGAQPGSQR LRHHAGNPLHLDVLVVDEASMIDLPMMSRLIDALPDHARVIF LGDRDQLASVEAGAVLGDICAYANAGFTAERARQLSRLTGT HVPAGTGTEAASLRDSLCLLQKSYRFGSDSGIGQLAAAINRG DKTAVKTVFQQDFTDIEKRLLQSGEDYIAMLEEALAGYGRYL DLLQARAEPDLIIQAFNEYQLLCALREGPFGVAGLNERIEQFM QQKRKIHRHPHSRWYEGRPVMIARNDSALGLFNGDIGIALDR GQGTRVWFAMPDGNIKSVQPSRLPEHETTWAMTVHKSQGSE FDHAALILPSQRTPVVTRELVYTAVTRARRRLSLYADERILSA AIATRTERRSGLAALFSSRE (SEQ ID NO:41)
RecQ	*E. coli*	MAQAEVLNLESGAKQVLQETFGYQQFRPGQEEIIDTVLSGRD CLVVMPTGGGKSLCYQIPALLLNGLTVVVSPLISLMKDQVDQ LQANGVAAACLNSTQTREQQLEVMTGCRTGQIRLLYIAPERL MLDNFLEHLAHWNPVLLAVDEAHCISQWGHDFRPEYAALGQ LRQRFPTLPFMALTATADDTTRQDIVRLLGLNDPLIQISSFDRP NIRYMLMEKFKPLDQLMRYVQEQRGKSGIIYCNSRAKVEDTA ARLQSKGISAAAYHAGLENNVRADVQEKFQRDDLQIVVATV AFGMGINKPNVRFVVHFDIPRNIESYYQETGRAGRDGLPAEA MLFYDPADMAWLRRCLEEKPQGQLQDIERHKLNAMGAFAE AQTCRRLVLLNYFGEGRQEPCGNCDICLDPPKQYDGSTDAQI ALSTIGRVNQRFGMGYVVEVIRGANNQRIRDYGHDKLKVYG MGRDKSHEHWVSVIRQLIHLGLVTQNIAQHSALQLTEAARPV LRGESSLQLAVPRIVALKPKAMQKSFGGNYDRKLFAKLRKLR KSIADESNVPPYVVFNDATLIEMAEQMPITASEMLSVNGVGM RKLERFGKPFMALIRAHVDGDDEE (SEQ ID NO:42)
Rep_Delta_ 2B	*E. coli*	MRLNPGQQQAVEFVTGPCLVLAGAGSGKTRVITNKIAHLIRG CGYQARHIAAVTFTNKAAREMKERVGQTLGRKEARGLMIST FHTLGLDIIKREYAALGMKANFSLFDDTDQLALLKELTEGLIE DDKVLLQQLISTISNWKNDLKTPAQAAAEAKGERDRIFAHCY GLYDAHLKACNVLDFDDLILLPTLLLQRNEEVRERWQNKIRY LLVDEYQDTNTSQYELVKLLVGSRARFTVVGDDDQSIYSWR GARPQNLVLLSQDFPALKVIKLEQNYRSSGRILKAANILIANN PHVFEKRLFSELGYGTELKVLSANNEEHEAERVTGELIAHHFV NKTQYKDYAILYRGNHQSRVFEKFLMQNRIPYKISGGGGGGE SEEELDQVQLMTLHASKGLEFPYVYMVGMEEGFLPHQSSIDE DNIDEERRLAYVGITRAQKELTFTLCKERRQYGELVRPEPSRF LLELPQDDLIWEQERKVVSAEERMQKGQSHLANLKAMMAA KRGK (SEQ ID NO:43)
UvrD	*E. coli*	MDVSYLLDSLNDKQREAVAAPRSNLLVLAGAGSGKTRVLVH RIAWLMSVENCSPYSIMAVTFTNKAAAEMRHRIGQLMGTSQ GGMWVGTFHGLAHRLLRAHHMDANLPQDFQILDSEDQLRLL KRLIKAMNLDEKQWPPRQAMWYINSQKDEGLRPHHIQSYGN PVEQTWQKVYQAYQEACDRAGLVDFAELLLRAHELWLNKP HILQHYRERFTNILVDEFQDTNNIQYAWIRLLAGDTGKVMIV GDDDQSIYGWRGAQVENIQRFLNDFPGAETIRLEQNYRSTSNI LSAANALIENNNGRLGKKLWTDGADGEPISLYCAFNELDEAR FVVNRIKTWQDNGGALAECAILYRSNAQSRVLEEALLQASMP

		YRIYGGMRFFERQEIKDALSYLRLIANRNDDAAFERVVNTPTR GIGDRTLDVVRQTSRDRQLTLWQACRELLQEKALAGRAASA LQRFMELIDALAQETADMPLHVQTDRVIKDSGLRTMYEQEK GEKGQTRIENLEELVTATRQFSYNEEDEDLMPLQAFLSHAAL EAGEGQADTWQDAVQLMTLHSAKGLEFPQVFIVGMEEGMFP SQMSLDEGGRLEEERRLAYVGVTRAMQKLTLTYAETRRLYG KEVYHRPSRFIGELPEECVEEVRLRATVSRPVSHQRMGTPMV ENDSGYKLGQRVRHAKFGEGTIVNMEGSGEHSRLQVAFQGQ GIKWLVAAYARLESV (SEQ ID NO:44)
RAPA	*E. coli*	MPFTLGQRWISDTESELGLGTVVAVDARTVTLLFPSTGENRL YARSDSPVTRVMFNPGDTITSHDGWQMQVEEVKEENGLLTYI GTRLDTEESGVALREVFLDSKLVFSKPQDRLFAGQIDRMDRF ALRYRARKYSSEQFRMPYSGLRGQRTSLIPHQLNIAHDVGRR HAPRVLLADEVGLGKTIEAGMILHQQLLSGAAERVLIIVPETL QHQWLVEMLRRFNLRFALFDDERYAEAQHDAYNPFDTEQLV ICSLDFARRSKQRLEHLCEAEWDLLVVDEAHHLVWSEDAPSR EYQAIEQLAEHVPGVLLLTATPEQLGMESHFARLRLLDPNRF HDFAQFVEEQKNYRPVADAVAMLLAGNKLSNDELNMLGEM IGEQDIEPLLQAANSDSEDAQSARQELVSMLMDRHGTSRVLF RNTRNGVKGFPKRELHTIKLPLPTQYQTAIKVSGIMGARKSAE DRARDMLYPERIYQEFEGDNATWWNFDPRVEWLMGYLTSH RSQKVLVICAKAATALQLEQVLREREGIRAAVFHEGMSIIERD RAAAWFAEEDTGAQVLLCSEIGSEGRNFQFASHMVMFDLPFN PDLLEQRIGRLDRIGQAHDIQIHVPYLEKTAQSVLVRWYHEGL DAFEHTCPTGRTIYDSVYNDLINYLASPDQTEGFDDLIKNCRE QHEALKAQLEQGRDRLLEIHSNGGEKAQALAESIEEQDDDTN LIAFAMNLFDIIGINQDDRGDNMIVLTPSDHMLVPDFPGLSED GITITFDREVALAREDAQFITWEHPLIRNGLDLILSGDTGSSTIS LLKNKALPVGTLLVELIYVVEAQAPKQLQLNRFLPPTPVRML LDKNGNNLAAQVEFETFNRQLNAVNRHTGSKLVNAVQQDV HAILQLGEAQIEKSARALIDAARNEADEKLSAELSRLEALRAV NPNIRDDELTAIESNRQQVMESLDQAGWRLDALRLIVVTHQ (SEQ ID NO:45)
RECG	*E. coli*	MKGRLLDAVPLSSLTGVGAALSNKLAKINLHTVQDLLLHLPL RYEDRTHLYPIGELLPGVYATVEGEVLNCNISFGGRRMMTCQ ISDGSGILTMRFFNFSAAMKNSLAAGRRVLAYGEAKRGKYGA EMIHPEYRVQGDLSTPELQETLTPVYPTTEGVKQATLRKLTD QALDLLDTCAIEELLPPELSQGMMTLPEALRTLHRPPPTLQLS DLETGQHPAQRRLILEELLAHNLSMLALRAGAQRFHAQPLSA NDTLKNKLLAALPFKPTGAQARVVAEIERDMALDVPMMRLV QGDVGSGKTLVAALAALRAIAHGKQVALMAPTELLAEQHAN NFRNWFAPLGIEVGWLAGKQKGKARLAQQEAIASGQVQMIV GTHAIFQEQVQFNGLALVIIDEQHRFGVHQRLALWEKGQQQG FHPHQLIMTATPIPRTLAMTAYADLDTSVIDELPPGRTPVTTV AIPDTRRTDIIDRVHHACITEGRQAYWVCTLIEESELLEAQAAE ATWEELKLALPELNVGLVHGRMKPAEKQAVMASFKQGELH LLVATTVIEVGVDVPNASLMIIENPERLGLAQLHQLRGRVGR GAVASHCVLLYKTPLSKTAQIRLQVLRDSNDGFVIAQKDLEIR GPGELLGTRQTGNAEFKVADLLRDQAMIPEVQRLARHIHERY PQQAKALIERWMPETERYSNA (SEQ ID NO:46)
DnaA	*E. coli*	MSLSLWQQCLARLQDELPATEFSMWIRPLQAELSDNTLALYA PNRFVLDWVRDKYLNNINGLLTSFCGADAPQLRFEVGTKPVT QTPQAAVTSNVAAPAQVAQTQPQRAAPSTRSGWDNVPAPAE PTYRSNVNVKHTFDNFVEGKSNQLARAAARQVADNPGGAY

		NPLFLYGGTGLGKTHLLHAVGNGIMARKPNAKVVYMHSERF VQDMVKALQNNAIEEFKRYYRSVDALLIDDIQFFANKERSQE EFFHTFNALLEGNQQIILTSDRYPKEINGVEDRLKSRFGWGLT VAIEPPELETRVAILMKKADENDIRLPGEVAFFIAKRLRSNVRE LEGALNRVIANANFTGRAITIDFVREALRDLLALQEKLVTIDNI QKTVAEYYKIKVADLLSKRRSRSVARPRQMAMALAKELTNH SLPEIGDAFGGRDHTTVLHACRKIEQLREESHDIKEDFSNLIRT LSS (SEQ ID NO:47)
HMGB1	*H. sapiens*	MGKGDPKKPRGKMSSYAFFVQTCREEHKKKHPDASVNFSEF SKKCSERWKTMSAKEKGKFEDMAKADKARYEREMKTYIPPK GETKKKFKDPNAPKRPPSAFFLFCSEYRPKIKGEHPGLSIGDV AKKLGEMWNNTAADDKQPYEKKAAKLKEKYEKDIAAYRAK GKPDAAKKGVVKAEKSKKKKEEEEDEEDEEDEEEEEDEEDE DEEEDDDDE (SEQ ID NO:48)
HMGB2	*H. sapiens*	MGKGDPNKPRGKMSSYAFFVQTCREEHKKKHPDSSVNFAEF SKKCSERWKTMSAKEKSKFEDMAKSDKARYDREMKNYVPP KGDKKGKKKDPNAPKRPPSAFFLFCSEHRPKIKSEHPGLSIGD TAKKLGEMWSEQSAKDKQPYEQKAAKLKEKYEKDIAAYRA KGKSEAGKKGPGRPTGSKKKNEPEDEEEEEEEEDEDEEEEDE DEE (SEQ ID NO:49)
RFA1	*H. sapiens*	MVGQLSEGAIAAIMQKGDTNIKPILQVINIRPITTGNSPPRYRL LMSDGLNTLSSFMLATQLNPLVEEEQLSSNCVCQIHRFIVNTL KDGRRVVILMELEVLKSAEAVGVKIGNPVPYNEGLGQPQVAP PAPAASPAASSRPQPQNGSSGMGSTVSKAYGASKTFGKAAGP SLSHTSGGTQSKVVPIASLTPYQSKWTICARVTNKSQIRTWSN SRGEGKLFSLELVDESGEIRATAFNEQVDKFFPLIEVNKVYYF SKGTLKIANKQFTAVKNDYEMTFNNETSVMPCEDDHHLPTV QFDFTGIDDLENKSKDSLVDIIGICKSYEDATKITVRSNNREVA KRNIYLMDTSGKVVTATLWGEDADKFDGSRQPVLAIKGARV SDFGGRSLSVLSSSTIIANPDIPEAYKLRGWFDAEGQALDGVSI SDLKSGGVGGSNTNWKTLYEVKSENLGQGDKPDYFSSVATV VYLRKENCMYQACPTQDCNKKVIDQQNGLYRCEKCDTEFPN FKYRMILSVNIADFQENQWVTCFQESAEAILGQNAAYLGELK DKNEQAFEEVFQNANFRSFIFRVRVKVETYNDESRIKATVMD VKPVDYREYGRRLVMSIRRSALM (SEQ ID NO:50)
RFA2	*H. sapiens*	MWNSGFESYGSSSYGGAGGYTQSPGGFGSPAPSQAEKKSRAR AQHIVPCTISQLLSATLVDEVFRIGNVEISQVTIVGIIRHAEKAP TNIVYKIDDMTAAPMDVRQWVDTDDTSSENTVVPPETYVKV AGHLRSFQNKKSLVAFKIMPLEDMNEFTTHILEVINAHMVLS KANSQPSAGRAPISNPGMSEAGNFGGNSFMPANGLTVAQNQ VLNLIKACPRPEGLNFQDLKNQLKHMSVSSIKQAVDFLSNEG HIYSTVDDDHFKSTDAE (SEQ ID NO:51)
RFA3	*H. sapiens*	MVDMMDLPRSRINAGMLAQFIDKPVCFVGRLEKIHPTGKMFI LSDGEGKNGTIELMEPLDEEISGIVEVVGRVTAKATILCTSYV QFKEDSHPFDLGLYNEAVKIIHDFPQFYPLGIVQHD (SEQ ID NO:52)
RAD51	*H. sapiens*	MAMQMQLEANADTSVEEESFGPQPISRLEQCGINANDVKKLE EAGFHTVEAVAYAPKKELINIKGISEAKADKILAEAAKLVPM GFTTATEFHQRRSEIIQITTGSKELDKLLQGGIETGSITEMFGEF RTGKTQICHTLAVTCQLPIDRGGGEGKAMYIDTEGTFRPERLL AVAERYGLSGSDVLDNVAYARAFNTDHQTQLLYQASAMMV ESRYALLIVDSATALYRTDYSGRGELSARQMHLARFLRMLLR LADEFGVAVVITNQVVAQVDGAAMFAADPKKPIGGNIIAHAS

		TTRLYLRKGRGETRICKIYDSPCLPEAEAMFAINADGVGDAK D (SEQ ID NO:53)
RAD52	*H. sapiens*	MSGTEEAILGGRDSHPAAGGGSVLCFGQCQYTAEEYQAIQKA LRQRLGPEYISSRMAGGGQKVCYIEGHRVINLANEMFGYNG WAHSITQQNVDFVDLNNGKFYVGVCAFVRVQLKDGSYHED VGYGVSEGLKSKALSLEKARKEAVTDGLKRALRSFGNALGN CILDKDYLRSLNKLPRQLPLEVDLTKAKRQDLEPSVEEARYNS CRPNMALGHPQLQQVTSPSRPSHAVIPADQDCSSRSLSSSAVE SEATHQRKLRQKQLQQQFRERMEKQQVRVSTPSAEKSEAAPP APPVTHSTPVTVSEPLLEKDFLAGVTQELIKTLEDNSEKWAVT PDAGDGVVKPSSRADPAQTSDTLALNNQMVTQNRTPHSVCH QKPQAKSGSWDLQTYSADQRTTGNWESHRKSQDMKKRKYD PS (SEQ ID NO:54)
BLM	*H. sapiens*	MAAVPQNNLQEQLERHSARTLNNKLSLSKPKFSGFTFKKKTS SDNNVSVTNVSVAKTPVLRNKDVNVTEDFSFSEPLPNTTNQQ RVKDFFKNAPAGQETQRGGSKSLLPDFLQTPKEVVCTTQNTP TVKKSRDTALKKLEFSSSPDSLSTINDWDDMDDFDTSETSKSF VTPPQSHFVRVSTAQKSKKGKRNFFKAQLYTTNTVKTDLPPP SSESEQIDLTEEQKDDSEWLSSDVICIDDGPIAEVHINEDAQES DSLKTHLEDERDNSEKKKNLEEAELHSTEKVPCIEFDDDDYD TDFVPPSPEEIISASSSSSKCLSTLKDLDTSDRKEDVLSTSKDLL SKPEKMSMQELNPETSTDCDARQISLQQQLIHVMEHICKLIDTI PDDKLKLLDCGNELLQQRNIRRKLLTEVDFNKSDASLLGSLW RYRPDSLDGPMEGDSCPTGNSMKELNFSHLPSNSVSPGDCLL TTTLGKTGFSATRKNLFERPLFNTHLQKSFVSSNWAETPRLGK KNESSYFPGNVLTSTAVKDQNKHTASINDLERETQPSYDIDNF DIDDFDDDDDWEDIMHNLAASKSSTAAYQPIKEGRPIKSVSER LSSAKTDCLPVSSTAQNINFSESIQNYTDKSAQNLASRNLKHE RFQSLSFPHTKEMMKIFHKKFGLHNFRTNQLEAINAALLGED CFILMPTGGGKSLCYQLPACVSPGVTVVISPLRSLIVDQVQKL TSLDIPATYLTGDKTDSEATNIYLQLSKKDPIIKLLYVTPEKICA SNRLISTLENLYERKLLARFVIDEAHCVSQWGHDFRQDYKRM NMLRQKFPSVPVMALTATANPRVQKDILTQLKILRPQVFSMS FNRHNLKYYVLPKKPKKVAFDCLEWIRKHHPYDSGIIYCLSR RECDTMADTLQRDGLAALAYHAGLSDSARDEVQQKWINQD GCQVICATIAFGMGIDKPDVRFVIHASLPKSVEGYYQESGRAG RDGEISHCLLFYTYHDVTRLKRLIMMEKDGNHHTRETHFNNL YSMVHYCENITECRRIQLLAYFGENGFNPDFCKKHPDVSCDN CCKTKDYKTRDVTDDVKSIVRFVQEHSSSQGMRNIKHVGPSG RFTMNMLVDIFLGSKSAKIQSGIFGKGSAYSRHNAERLFKKLI LDKILDEDLYINANDQAIAYVMLGNKAQTVLNGNLKVDFME TENSSSVKKQKALVAKVSQREEMVKKCLGELTEVCKSLGKV FGVHYFNIFNTVTLKKLAESLSSDPEVLLQIDGVTEDKLEKYG AEVISVLQKYSEWTSPAEDSSPGISLSSSRGPGRSAAEELDEEIP VSSHYFASKTRNERKRKKMPASQRSKRRKTASSGSKAKGGS ATCRKISSKTKSSSIIGSSSASHTSQATSGANSKLGIMAPPKPIN RPFLKPSYAFS (SEQ ID NO:55)
EXO1	*H. sapiens*	MGIQGLLQFIKEASEPIHVRKYKGQVVAVDTYCWLHKGAIAC AEKLAKGEPTDRYVGFCMKFVNMLLSHGIKPILVFDGCTLPS KKEVERSRRERRQANLLKGKQLLREGKVSEARECFTRSINITH AMAHKVIKAARSQGVDCLVAPYEADAQLAYLNKAGIVQAIIT EDSDLLAFGCKKVILKMDQFGNGLEIDQARLGMCRQLGDVF TEEKFRYMCILSGCDYLSSLRGIGLAKACKVLRLANNPDIVKV IKKIGHYLKMNITVPEDYINGFIRANNTFLYQLVFDPIKRKLIPL

		NAYEDDVDPETLSYAGQYVDDSIALQIALGNKDINTFEQIDDY NPDTAMPAHSRSHSWDDKTCQKSANVSSIWHRNYSPRPESGT VSDAPQLKENPSTVGVERVISTKGLNLPRKSSIVKRPRSAELSE DDLLSQYSLSFTKKTKKNSSEGNKSLSFSEVFVPDLVNGPTNK KSVSTPPRTRNKFATFLQRKNEESGAVVVPGTRSRFFCSSDST DCVSNKVSIQPLDETAVTDKENNLHESEYGDQEGKRLVDTDV ARNSSDDIPNNHIPGDHIPDKATVFTDEESYSFESSKFTRTISPP TLGTLRSCFSWSGGLGDFSRTPSPSPSTALQQFRRKSDSPTSLP ENNMSDVSQLKSEESSDDESHPLREEACSSQSQESGEFSLQSS NASKLSQCSSKDSDSEESDCNIKLLDSQSDQTSKLRLSHFSKK DTPLRNKVPGLYKSSSADSLSTTKIKPLGPARASGLSKKPASIQ KRKHHNAENKPGLQIKLNELWKNFGFKKDSEKLPPCKKPLSP VRDNIQLTPEAEEDIFNKPECGRVQRAIFQ (SEQ ID NO:56)
RECQ1	*H. sapiens*	MASVSALTEELDSITSELHAVEIQIQELTERQQELIQKKKVLTK KIKQCLEDSDAGASNEYDSSPAAWNKEDFPWSGKVKDILQN VFKLEKFRPLQLETINVTMAGKEVFLVMPTGGGKSLCYQLPA LCSDGFTLVICPLISLMEDQLMVLKQLGISATMLNASSSKEHV KWVHAEMVNKNSELKLIYVTPEKIAKSKMFMSRLEKAYEAR RFTRIAVDEVHCCSQWGHDFRPDYKALGILKRQFPNASLIGLT ATATNHVLTDAQKILCIEKCFTFTASFNRPNLYYEVRQKPSNT EDFIEDIVKLINGRYKGQSGIIYCFSQKDSEQVTVSLQNLGIHA GAYHANLEPEDKTTVHRKWSANEIQVVVATVAFGMGIDKPD VRFVIHHSMSKSMENYYQESGRAGRDDMKADCILYYGFGDIF RISSMVVMENVGQQKLYEMVSYCQNISKCRRVLMAQHFDEV WNSEACNKMCDNCCKDSAFERKNITEYCRDLIKILKQAEELN EKLTPLKLIDSWMGKGAAKLRVAGVVAPTLPREDLEKIIAHF LIQQYLKEDYSFTAYATISYLKIGPKANLLNNEAHAITMQVTK STQNSFRAESSQTCHSEQGDKKMEEKNSGNFQKKAANMLQQ SGSKNTGAKKRKIDDA (SEQ ID NO:57)
ERCC2	*H. sapiens*	MKLNVDGLLVYFPYDYIYPEQFSYMRELKRTLDAKGHGVLE MPSGTGKTVSLLALIMAYQRAYPLEVTKLIYCSRTVPEIEKVIE ELRKLLNFYEKQEGEKLPFLGLALSSRKNLCIHPEVTPLRFGK DVDGKCHSLTASYVRAQYQHDTSLPHCRFYEEFDAHGREVP LPAGIYNLDDLKALGRRQGWCPYFLARYSILHANVVVYSYH YLLDPKIADLVSKELARKAVVVFDEAHNIDNVCIDSMSVNLT RRTLDRCQGNLETLQKTVLRIKETDEQRLRDEYRRLVEGLRE ASAARETDAHLANPVLPDEVLQEAVPGSIRTAEHFLGFLRRLL EYVKWRLRVQHVVQESPPAFLSGLAQRVCIQRKPLRFCAERL RSLLHTLEITDLADFSPLTLLANFATLVSTYAKGFTIIEPFDDR TPTIANPILHFSCMDASLAIKPVFERFQSVIITSGTLSPLDIYPKI LDFHPVTMATFTMTLARVCLCPMIIGRGNDQVAISSKFETRED IAVIRNYGNLLLEMSAVVPDGIVAFFTSYQYMESTVASWYEQ GILENIQRNKLLFIETQDGAETSVALEKYQEACENGRGAILLS VARGKVSEGIDFVHHYGRAVIMFGVPYVYTQSRILKARLEYL RDQFQIRENDFLTFDAMRHAAQCVGRAIRGKTDYGLMVFAD KRFARGDKRGKLPRWIQEHLTDANLNLTVDEGVQVAKYFLR QMAQPFHREDQLGLSLLSLEQLESEETLKRIEQIAQQL (SEQ ID NO:58)
ERCC3	*H. sapiens*	MGKRDRADRDKKKSRKRHYEDEEDDEEDAPGNDPQEAVPS AAGKQVDESGTKVDEYGAKDYRLQMPLKDDHTSRPLWVAP DGHIFLEAFSPVYKYAQDFLVAIAEPVCRPTHVHEYKLTAYSL YAAVSVGLQTSDITEYLRKLSKTGVPDGIMQFIKLCTVSYGKV KLVLKHNRYFVESCHPDVIQHLLQDPVIRECRLRNSEGEATEL ITETFTSKSAISKTAESSGGPSTSRVTDPQGKSDIPMDLFDFYE

		QMDKDEEEEEETQTVSFEVKQEMIEELQKRCIHLEYPLLAEY DFRNDSVNPDINIDLKPTAVLRPYQEKSLRKMFGNGRARSGVI VLPCGAGKSLVGVTAACTVRKRCLVLGNSAVSVEQWKAQFK MWSTIDDSQICRFTSDAKDKPIGCSVAISTYSMLGHTTKRSWE AERVMEWLKTQEWGLMILDEVHTIPAKMFRRVLTIVQAHCK LGLTATLVREDDKIVDLNFLIGPKLYEANWMELQNNGYIAKV QCAEVWCPMSPEFYREYVAIKTKKRILLYTMNPNKFRACQFL IKFHERRNDKIIVFADNVFALKEYAIRLNKPYIYGPTSQGERM QILQNFKHNPKINTIFISKVGDTSFDLPEANVLIQISSHGGSRRQ EAQRLGRVLRAKKGMVAEEYNAFFYSLVSQDTQEMAYSTKR QRFLVDQGYSFKVITKLAGMEEEDLAFSTKEEQQQLLQKVLA ATDLDAEEEVVAGEFGSRSSQASRRFGTMSSMSGADDTVYM EYHSSRSKAPSKHVHPLFKRFRK (SEQ ID NO:59)
WRN	*H. sapiens*	MSEKKLETTAQQRKCPEWMNVQNKRCAVEERKACVRKSVF EDDLPFLEFTGSIVYSYDASDCSFLSEDISMSLSDGDVVGFDM EWPPLYNRGKLGKVALIQLCVSESKCYLFHVSSMSVFPQGLK MLLENKAVKKAGVGIEGDQWKLLRDFDIKLKNFVELTDVAN KKLKCTETWSLNSLVKHLLGKQLLKDKSIRCSNWSKFPLTED QKLYAATDAYAGFIIYRNLEILDDTVQRFAINKEEEILLSDMN KQLTSISEEVMDLAKHLPHAFSKLENPRRVSILLKDISENLYSL RRMIIGSTNIETELRPSNNLNLLSFEDSTTGGVQQKQIREHEVLI HVEDETWDPTLDHLAKHDGEDVLGNKVERKEDGFEDGVED NKLKENMERACLMSLDITEHELQILEQQSQEEYLSDIAYKSTE HLSPNDNENDTSYVIESDEDLEMEMLKHLSPNDNENDTSYVI ESDEDLEMEMLKSLENLNSGTVEPTHSKCLKMERNLGLPTKE EEEDDENEANEGEEDDDKDFLWPAPNEEQVTCLKMYFGHSS FKPVQWKVIHSVLEERRDNVAVMATGYGKSLCFQYPPVYVG KIGLVISPLISLMEDQVLQLKMSNIPACFLGSAQSENVLTDIKL GKYRIVYVTPEYCSGNMGLLQQLEADIGITLIAVDEAHCISEW GHDFRDSFRKLGSLKTALPMVPIVALTATASSSIREDIVRCLNL RNPQITCTGFDRPNLYLEVRRKTGNILQDLQPFLVKTSSHWEF EGPTIIYCPSRKMTQQVTGELRKLNLSCGTYHAGMSFSTRKDI HHRFVRDEIQCVIATIAFGMGINKADIRQVIHYGAPKDMESYY QEIGRAGRDGLQSSCHVLWAPADINLNRHLLTEIRNEKFRLY KLKMMAKMEKYLHSSRCRRQIILSHFEDKQVQKASLGIMGTE KCCDNCRSRLDHCYSMDDSEDTSWDFGPQAFKLLSAVDILGE KFGIGLPILFLRGSNSQRLADQYRRHSLFGTGKDQTESWWKA FSRQLITEGFLVEVSRYNKFMKICALTKKGRNWLHKANTESQ SLILQANEELCPKKLLLPSSKTVSSGTKEHCYNQVPVELSTEK KSNLEKLYSYKPCDKISSGSNISKKSIMVQSPEKAYSSSQPVIS AQEQETQIVLYGKLVEARQKHANKMDVPPAILATNKILVDM AKMRPTTVENVKRIDGVSEGKAAMLAPLLEVIKHFCQTNSVQ TDLFSSTKPQEEQKTSLVAKNKICTLSQSMAITYSLFQEKKMP LKSIAESRILPLMTIGMHLSQAVKAGCPLDLERAGLTPEVQKII ADVIRNPPVNSDMSKISLIRMLVPENIDTYLIHMAIEILKHGPD SGLQPSCDVNKRRCFPGSEEICSSSKRSKEEVGINTETSSAERK RRLPVWFAKGSDTSKKLMDKTKRGGLFS (SEQ ID NO:60)
RECQ5	*H. sapiens*	MSSHHTTFPFDPERRVRSTLKKVFGFDSFKTPLQESATMAVV KGNKDVFVCMPTGAGKSLCYQLPALLAKGITIVVSPLIALIQD QVDHLLTLKVRVSSLNSKLSAQERKELLADLEREKPQTKILYI TPEMAASSSFQPTLNSLVSRHLLSYLVVDEAHCVSQWGHDFR PDYLRLGALRSRLGHAPCVALTATATPQVQEDVFAALHLKKP VAIFKTPCFRANLFYDVQFKELISDPYGNLKDFCLKALGQEAD KGLSGCGIVYCRTREACEQLAIELSCRGVNAKAYHAGLKASE RTLVQNDWMEEKVPVIVATISFGMGVDKANVRFVAHWNIAK

		SMAGYYQESGRAGRDGKPSWCRLYYSRNDRDQVSFLIRKEV AKLQEKRGNKASDKATIMAFDALVTFCEELGCRHAAIAKYFG DALPACAKGCDHCQNPTAVRRRLEALERSSSWSKTCIGPSQG NGFDPELYEGGRKGYGDFSRYDEGSGGSGDEGRDEAHKREW NLFYQKQMQLRKGKDPKIEEFVPPDENCPLKEASSRRIPRLTV KAREHCLRLLEEALSSNRQSTRTADEADLRAKAVELEHETFR NAKVANLYKASVLKKVADIHRASKDGQPYDMGGSAKSCSA QAEPPEPNEYDIPPASHVYSLKPKRVGAGFPKGSCPFQTATEL METTRIREQAPQPERGGEHEPPSRPCGLLDEDGSEPLPGPRGE VPGGSAHYGGPSPEKKAKSSSGGSSLAKGRASKKQQLLATAA HKDSQSIARFFCRRVESPALLASAPEAEGACPSCEGVQGPPMA PEKYTGEEDGAGGHSPAPPQTEECLRERPSTCPPRDQGTPEVQ PTPAKDTWKGKRPRSQQENPESQPQKRPRPSAKPSVVAEVKG SVSASEQGTLNPTAQDPFQLSAPGVSLKEAANVVVKCLTPFY KEGKFASKELFKGFARHLSHLLTQKTSPGRSVKEEAQNLIRHF FHGRARCESEADWHGLCGPQR (SEQ ID NO:61)
RECQ4	*H. sapiens*	MERLRDVRERLQAWERAFRRQRGRRPSQDDVEAAPEETRAL YREYRTLKRTTGQAGGGLRSSESLPAAAEEAPEPRCWGPHLN RAATKSPQSTPGRSRQGSVPDYGQRLKANLKGTLQAGPALGR RPWPLGRASSKASTPKPPGTGPVPSFAEKVSDEPPQLPEPQPRP GRLQHLQASLSQRLGSLDPGWLQRCHSEVPDFLGAPKACRPD LGSEESQLLIPGESAVLGPGAGSQGPEASAFQEVSIRVGSPQPS SSGGEKRRWNEEPWESPAQVQQESSQAGPPSEGAGAVAVEE DPPGEPVQAQPPQPCSSPSNPRYHGLSPSSQARAGKAEGTAPL HIFPRLARHDRGNYVRLNMKQKHYVRGRALRSRLLRKQAW KQKWRKKGECFGGGGATVTTKESCFLNEQFDHWAAQCPRP ASEEDTDAVGPEPLVPSPQPVPEVPSLDPTVLPLYSLGPSGQLA ETPAEVFQALEQLGHQAFRPGQERAVMRILSGISTLLVLPTGA GKSLCYQLPALLYSRRSPCLTLVVSPLLSLMDDQVSGLPPCLK AACIHSGMTRKQRESVLQKIRAAQVHVLMLTPEALVGAGGL PPAAQLPPVAFACIDEAHCLSQWSHNFRPCYLRVCKVLRERM GVHCFLGLTATATRRTASDVAQHLAVAEEPDLHGPAPVPTNL HLSVSMDRDTDQALLTLLQGKRFQNLDSIIIYCNRREDTERIA ALLRTCLHAAWVPGSGGRAPKTTAEAYHAGMCSRERRRVQR AFMQGQLRVVVATVAFGMGLDRPDVRAVLHLGLPPSFESYV QAVGRAGRDGQPAHCHLFLQPQGEDLRELRRHVHADSTDFL AVKRLVQRVFPACTCTCTRPPSEQEGAVGGERPVPKYPPQEA EQLSHQAAPGPRRVCMGHERALPIQLTVQALDMPEEAIETLL CYLELHPHHWLELLATTYTHCRLNCPGGPAQLQALAHRCPPL AVCLAQQLPEDPGQGSSSVEFDMVKLVDSMGWELASVRRAL CQLQWDHEPRTGVRRGTGVLVEFSELAFHLRSPGDLTAEEKD QICDFLYGRVQARERQALARLRRTFQAFHSVAFPSCGPCLEQ QDEERSTRLKDLLGRYFEEEEGQEPGGMEDAQGPEPGQARLQ DWEDQVRCDIRQFLSLRPEEKFSSRAVARIFHGIGSPCYPAQV YGQDRRFWRKYLHLSFHALVGLATEELLQVAR (SEQ ID NO:62)
HELQ	*H. sapiens*	MDECGSRIRRRVSLPKRNRPSLGCIFGAPTAAELVPGDEGKEE EEMVAENRRRKTAGVLPVEVQPLLLSDSPECLVLGGGDTNPD LLRHMPTDRGVGDQPNDSEVDMFGDYDSFTENSFIAQVDDLE QKYMQLPEHKKHATDFATENLCSESIKNKLSITTIGNLTELQT DKHTENQSGYEGVTIEPGADLLYDVPSSQAIYFENLQNSSNDL GDHSMKERDWKSSSHNTVNEELPHNCIEQPQQNDESSSKVRT SSDMNRRKSIKDHLKNAMTGNAKAQTPIFSRSKQLKDTLLSE EINVAKKTVESSSNDLGPFYSLPSKVRDLYAQFKGIEKLYEWQ HTCLTLNSVQERKNLIYSLPTSGGKTLVAEILMLQELLCCRKD

		VLMILPYVAIVQEKISGLSSFGIELGFFVEEYAGSKGRFPPTKR REKKSLYIATIEKGHSLVNSLIETGRIDSLGLVVVDELHMIGEG SRGATLEMTLAKILYTSKTTQIIGMSATLNNVEDLQKFLQAEY YTSQFRPVELKEYLKINDTIYEVDSKAENGMTFSRLLNYKYSD TLKKMDPDHLVALVTEVIPNYSCLVFCPSKKNCENVAEMICK FLSKEYLKHKEKEKCEVIKNLKNIGNGNLCPVLKRTIPFGVAY HHSGLTSDERKLLEEAYSTGVLCLFTCTSTLAAGVNLPARRVI LRAPYVAKEFLKRNQYKQMIGRAGRAGIDTIGESILILQEKDK QQVLELITKPLENCYSHLVQEFTKGIQTLFLSLIGLKIATNLDDI YHFMNGTFFGVQQKVLLKEKSLWEITVESLRYLTEKGLLQKD TIYKSEEEVQYNFHITKLGRASFKGTIDLAYCDILYRDLKKGL EGLVLESLLHLIYLTTPYDLVSQCNPDWMIYFRQFSQLSPAEQ NVAAILGVSESFIGKKASGQAIGKKVDKNVVNRLYLSFVLYT LLKETNIWTVSEKFNMPRGYIQNLLTGTASFSSCVLHFCEELE EFWVYRALLVELTKKLTYCVKAELIPLMEVTGVLEGRAKQL YSAGYKSLMHLANANPEVLVRTIDHLSRRQAKQIVSSAKMLL HEKAEALQEEVEELLRLPSDFPGAVASSTDKA (SEQ ID NO:63)
ICP8	*H. simplex*	METKPKTATTIKVPPGPLGYVYARACPSEGIELLALLSARSGD SDVAVAPLVVGLTVESGFEANVAVVVGSRTTGLGGTAVSLK LTPSHYSSSVYVFHGGRHLDPSTQAPNLTRLCERARRHFGFSD YTPRPGDLKHETTGEALCERLGLDPDRALLYLVVTEGFKEAV CINNTFLHLGGSDKVTIGGAEVHRIPVYPLQLFMPDFSRVIAEP FNANHRSIGEKFTYPLPFFNRPLNRLLFEAVVGPAAVALRCRN VDAVARAAAHLAFDENHEGAALPADITFTAFEASQGKTPRGG RDGGGKGAAGGFEQRLASVMAGDAALALESIVSMAVFDEPP TDISAWPLFEGQDTAAARANAVGAYLARAAGLVGAMVFSTN SALHLTEVDDAGPADPKDHSKPSFYRFFLVPGTHVAANPQVD REGHVVPGFEGRPTAPLVGGTQEFAGEHLAMLCGFSPALLAK MLFYLERCDGAVIVGRQEMDVFRYVADSNQTDVPCNLCTFD TRHACVHTTLMRLRARHPKFASAARGAIGVFGTMNSMYSDC DVLGNYAAFSALKRADGSETARTIMQETYRAATERVMAELE TLQYVDQAVPTAMGRLETIITNREALHTVVNNVRQVVDREV EQLMRNLVEGRNFKFRDGLGEANHAMSLTLDPYACGPCPLL QLLGRRSNLAVYQDLALSQCHGVFAGQSVEGRNFRNQFQPV LRRRVMDMFNNGFLSAKTLTVALSEGAAICAPSLTAGQTAPA ESSFEGDVARVTLGFPKELRVKSRVLFAGASANASEAAKARV ASLQSAYQKPDKRVDILLGPLGFLLKQFHAAIFPNGKPPGSNQ PNPQWFWTALQRNQLPARLLSREDIETIAFIKKFSLDYGAINFI NLAPNNVSELAMYYMANQILRYCDHSTYFINTLTAIIAGSRRP PSVQAAAAWSAQGGAGLEAGARALMDAVDAHPGAWTSMF ASCNLLRPVMAARPMVVLGLSISKYYGMAGNDRVFQAGNW ASLMGGKNACPLLIFDRTRKFVLACPRAGFVCAASSLGGGAH ESSLCEQLRGIISEGGAAVASSVFVATVKSLGPRTQQLQIEDW LALLEDEYLSEEMMELTARALERGNGEWSTDAALEVAHEAE ALVSQLGNAGEVFNFGDFGCEDDNATPFGGPGAPGPAFAGR KRAFHGDDPFGEGPPDKKGDLTLDML (SEQ ID NO:64)
RAD25	*S. cerevisiae*	MTDVEGYQPKSKGKIFPDMGESFFSSDEDSPATDAEIDENYD DNRETSEGRGERDTGAMVTGLKKPRKKTKSSRHTAADSSMN QMDAKDKALLQDTNSDIPADFVPDSVSGMFRSHDFSYLRLRP DHASRPLWISPSDGRIILESFSPLAEQAQDFLVTIAEPISRPSHIH EYKITAYSLYAAVSVGLETDDIISVLDRLSKVPVAESIINFIKGA TISYGKVKLVIKHNRYFVETTQADILQMLLNDSVIGPLRIDSD HQVQPPEDVLQQQLQQTAGKPATNVNPNDVEAVFSAVIGGD NEREEEDDDIDAVHSFEIANESVEVVKKRCQEIDYPVLEEYDF RNDHRNPDLDIDLKPSTQIRPYQEKSLSKMFGNGRARSGIIVLP

		CGAGKTLVGITAACTIKKSVIVLCTSSVSVMQWRQQFLQWCT LQPENCAVFTSDNKEMFQTESGLVVSTYSMVANTRNRSHDS QKVMDFLTGREWGFIILDEVHVVPAAMFRRVVSTIAAHAKLG LTATLVREDDKIGDLNFLIGPKLYEANWMELSQKGHIANVQC AEVWCPMTAEFYQEYLRETARKRMLLYIMNPTKFQACQFLIQ YHERRGDKIIVFSDNVYALQEYALKMGKPFIYGSTPQQERMN ILQNFQYNDQINTIFLSKVGDTSIDLPEATCLIQISSHYGSRRQE AQRLGRILRAKRRNDEGFNAFFYSLVSKDTQEMYYSTKRQAF LVDQGYAFKVITHLHGMENIPNLAYASPRERRELLQEVLLKN EEAAGIEVGDDADNSVGRGSNGHKRFKSKAVRGEGSLSGLA GGEDMAYMEYSTNKNKELKEHHPLIRKMYYKNLKK (SEQ ID NO:65)
RAD3	*S. cerevisiae*	MKFYIDDLPVLFPYPKIYPEQYNYMCDIKKTLDVGGNSILEMP SGTGKTVSLLSLTIAYQMHYPEHRKIIYCSRTMSEIEKALVELE NLMDYRTKELGYQEDFRGLGLTSRKNLCLHPEVSKERKGTV VDEKCRRMTNGQAKRKLEEDPEANVELCEYHENLYNIEVED YLPKGVFSFEKLLKYCEEKTLCPYFIVRRMISLCNIIIYSYHYLL DPKIAERVSNEVSKDSIVIFDEAHNIDNVCIESLSLDLTTDALR RATRGANALDERISEVRKVDSQKLQDEYEKLVQGLHSADILT DQEEPFVETPVLPQDLLTEAIPGNIRRAEHFVSFLKRLIEYLKT RMKVLHVISETPKSFLQHLKQLTFIERKPLRFCSERLSLLVRTL EVTEVEDFTALKDIATFATLISTYEEGFLLIIEPYEIENAAVPNPI MRFTCLDASIAIKPVFERFSSVIITSGTISPLDMYPRMLNFKTVL QKSYAMTLAKKSFLPMIITKGSDQVAISSRFEIRNDPSIVRNYG SMLVEFAKITPDGMVVFFPSYLYMESIVSMWQTMGILDEVW KHKLILVETPDAQETSLALETYRKACSNGRGAILLSVARGKVS EGIDFDHQYGRTVLMIGIPFQYTESRILKARLEFMRENYRIREN DFLSFDAMRHAAQCLGRVLRGKDDYGVMVLADRRFSRKRS QLPKWIAQGLSDADLNLSTDMAISNTKQFLRTMAQPTDPKDQ EGVSVWSYEDLIKHQNSRKDQGGFIENENKEGEQDEDEDEDI EMQ (SEQ ID NO:66)
Cas9_RecII	*S. pyogenes*	DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRV (SEQ ID NO:67)
Cas9_RecI_ RecII_RecI	*S. pyogenes*	DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIY HLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRV NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIF FDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNR EKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEV VDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNE LTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQL KRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGD (SEQ ID NO:68)

[0106] In certain embodiments, the base editor (*e.g.*, a Cas-free base editor) comprises one or more DNA oligonucleotides, one or more RNA oligonucleotides

and/or one or more synthetic oligonucleotides including one or more peptide nucleic acids (PNA), one or more locked nucleic acids (LNA) and/or one or more bridged nucleic acids (BNA). *See, e.g.,* Figure 1D. *See, e.g.,* Nielsen, *et al.* (1991) *Science* 254:1497-1500 and Bahal, *et al.* (2014) *Curr Gene Ther.* 14(5):331-342 regarding PNA; Moreno, *et al.* (2013) *Nucleic Acid Res.* 41(5):3257-3273 and Geny, *et al.* (2016) *Nucleic Acid Res.* 44(5):2007-2019 regarding LNA; and Rahman, *et al.* (2007) *Nucleosides Nucleotides Nucleic Acids* 26(10-12):1625-1628 regarding BNA.

[0107] In certain embodiments, the Cas9-free base editors of the invention comprise a ZFP-deaminase fusion protein and a ZFN nickase, and optionally one or more DNA-destabilizing factors. In certain embodiments, the DNA-destabilizing factor is a protein (*e.g.,* as shown in Table A) or an oligonucleotide (*e.g.,* an LNA, PNA or BNA). In other embodiments, the Cas9-free base editors comprise non-Cas9 CRISPR proteins with DNA destabilizing (unwinding) properties, including any Cas9 equivalents such as Cas12a (including a full length or truncated Cas12 protein). In other embodiments, the Cas9-free base editors do not comprise any elements from a CRISPR system. The one or more non-Cas9 DNA-destabilizing (unwinding) factor(s) (*e.g.,* proteins of Table A, LNAs, PNAs, BNAs, etc.) may be operably linked to any component of the base editor, for example either component of the ZFP-deaminase fusion protein and/or any of the components of the ZFN nickase.

[0108] In certain embodiments, the base editor comprises one or more nucleotide sequences, for example one or more DNA oligonucleotides, RNA oligonucleotides, peptide nucleic acids (PNAs), locked nucleic acids (LNAs) and/or bridged nucleic acids (BNAs), which can be used to provide a single stranded DNA substrate for base editors at the target site. This can be facilitated by e.g. duplex invasion, triplex invasion or a tail-clamp (Quijano, *et al.* (2017) *Yale J. Biol and Med.* 90:583-598; Pellestor and Paulasova (2004) *European J. Human Genetics* 12:694-700; Schleifman, *et al.* (2011) *Chem & Bio.* 18:1189-1198. The structure of the one or more nucleotide sequences of the base editor will vary in length; number and position of DNA and/or RNA and/or LNA and/or LNA and/or BNA bases; phosphorothioate bonds; other common modifications) of these oligonucleotides depending on the target sequence composition.

[0109] In certain embodiments, the base editor comprises one or more PNAs, for examplegamma PNAs containing miniPEG substitutions and the gamma position for enhanced binding, increased solubility and improved delivery (Bahal, *et al.* (2014)

Current Gene Ther. 14(5):331-342. In certain embodiments, the PNAs comprise one or more O indicates 8-amino-2,6-dioxaoctanoic acid linkers and/or one or more cytosines (C) or pseudoisocytosine residues. Optionally, one or more lysine (Lys) residues are included in the PNA, for example on the N- and/or C-terminals of the PNA sequence. In certain embodiments, 1, 2, 3, 4, 5 or more Lys residues are included at one or both terminals of the PNA. In certain embodiments, two or more PNAs are used in the base editor, for example in the same or in reverse orientation relative to each other. In certain embodiments, the one or more PNAs have the structure: N-Lys-Lys-Lys-NNNNNNNNNN-OOO-NNNNNNNNNN-Lys-Lys-Lys-C; N-Lys-Lys-Lys-NNNNNNNNNN-OOO-NNNNNNNNNNNNNNN-Lys-Lys-Lys-C; N-Lys-Lys-Lys-NNNNNNNNNN-OOO-NNNNNNNNNN-Lys-Lys-Lys-C; N-Lys-Lys-Lys-NNNNNNNNNN-OOO-NNNNNNNNNN-Lys-Lys-Lys-C; and/or N-Lys-Lys-Lys-NNNNNNNNNNNNNNN-Lys-Lys-Lys-C, wherein O indicates 8-amino-2,6-dioxaoctanoic acid linkers and C indicates cytosine. The Lys resides on the N- and/or C-terminals of the PNA sequence are optional and pseudoisocytosine be can substituted for cytosine. In the certain embodiments, the one or more PNAs comprises one or more PNAs as shown in Figures 8B to 8E.

[0110] In other embodiments the base editor comprises one or more LNAs. LNAs can include a stacking linker and 2'-glycylamino-LNA for improved performance (Geny, *et al.* (2016) *Nucleic Acids Res.* 44(5):2007-2019. In certain embodiments, the LNA comprise one or more phosphorothioate bonds, optionally between one or more LNA residues and/or DNA residues. In other embodiments, the LNA comprises one or more Cholesterol-TEG, which may increase uptake into cells. In certain embodiments, the one or more LNAs have the following structure: 5'-NnNnNnNnNnNnNnNtctctnNnNnNnNnNnNnNnnNnnNnnNnnNn-3' (SEQ ID NO:1); 5'-N*n*NnNnNnNnNnNnNtctctnNnNnNnNnNnNnNnnNnnNnnNnn*N*n-3' (SEQ ID NO:69); and/or 5'-NnNnNnNnNnNnNnNtctctnNnNnNnNnNnNnNnnNnnNnnNnnNn-Chol-TEG-3' (SEQ ID NO:70), where LNA nucleotides are in uppercase; DNA nucleotides are in lower case; "*" indicates phosphorothioate bonds; and "Chol-TEG" indicates 3' Cholesterol-TEG (*see, e.g.*, Bijsterbosch, *et al.* (2000) *Nucleic Acids Res.* 28:2717-2725; Bijsterbosch, *et al.* (2002) *J. Pharmacol. Exp. Ther.* 302:619-626; Manoharan (2002) *Antisense Nucleic Acid Drug Dev* 12:103-28; M. Manoharan (2004) *Curr Opin*

Chem Biol. 8:570-9*)* for increased uptake into cells. In certain embodiments, the base editor comprises one or more LNAs as shown in Figure 8F or Figure 8G.

[0111] The one or more DNA destabilizing factors may be provided independently from and/or with a base editor (*e.g.*, ABE and/or CBE) and/or a nickase. In certain embodiments, the DNA de-stabilizing factor(s) is(are) fused to the ZFP and/or ZFN nickase in any orientation (*e.g.*, N- and/or C-terminal). The DNA destabilizing factor(s) can bind within a 1kb window of the base editor target site.

[0112] In certain embodiments, described herein is a base editing system comprising a base editor (*e.g.*, a Cas9 adenine or cytosine base editor) and one or more additional DNA-binding domains (*e.g.*, ZFPs, TALEs, additional sgRNA) that specifically binds to a target site near (RANGE) the base editor. In certain embodiments, the Cas9-containing base editing system comprises: a Cas9 nickase and one or more DNA-binding molecules (*e.g.*, ZFPs) that serve to anchor the Cas9 nickase and/or position one or more components of the base editor in relation to the other components, thereby increasing specificity and/or efficiency of base editing. The one or more DNA-binding molecules (*e.g.*, ZFP anchor(s)) typically bind to a target site within 1-50 (or any value therebetween) nucleotides of the base editor (*e.g.*, Cas9 or Cas9-free base editor) and/or targeted base. The DNA-binding molecule may bind 5' and/or 3' to the base editor and/or targeted base. In certain embodiments, the Cas9 base editor as described herein comprises two more ZFP domains, for example a ZFP domain operably linked to a deaminase domain or component thereof and a ZFP anchor domain. *See, e.g.*, Figure 1B. In certain embodiments, the at least one DNA-binding domain binds to a target site 5' of the base editor, optionally on the same or different strand as bound by the base editor. *See, e.g.*, Figure 1B, bottom right schematic; Figure 3; and Figure 7A for exemplary embodiments in which one or more additional ZFP anchors are used and specifically binds to a target site 5' of the base editor of the system and on the same strand as bound by the base editor. The inclusion of one or more ZFP anchors can increase efficiency and/or specificity of the base editor, for example in some cases, 2-fold to 5-fold (or any value therebetween), 10-fold to 100-fold (or any value therebetween), or more than 100-fold as compared to base editors not including a ZFP anchor.

[0113] In other embodiments, described herein is a base editing system comprising: (1) a Cas9 nickase (*e.g.*, comprising a catalytically inactive monomer of an adenosine deaminase); (2) an anchor DNA-binding domain (*e.g.*, ZFP) that

specifically binds to a target site 5' or 3' to the Cas9 nickase; and (3) a non-Cas9 nickase (*e.g.*, a ZFP-nickase) comprising a catalytically inactive monomer of the adenosine deaminase and a DNA-binding domain (*e.g.*, ZFP) that binds to a target site 5' or 3' to the Cas9 nickase. Upon dimerization of the A deaminase monomers of the Cas9 nickase and non-Cas9 nickase (*e.g.*, ZFN nickase) dimerize form a functional deaminase. The anchor DNA-binding domain and non-Cas9 nickase may bind on the same or different strands of the target and/or on the same (5' or 3') or different (one on 5' and one on 3') sides of the Cas9 nickase. In certain embodiments, the anchor DNA-binding domain and the non-Cas9 nickase bind to opposite strands on opposite sides of the Cas9 nickase. *See, e.g.*, Figure 1B, top schematics.

[0114] In other embodiments, described herein is a base editing system comprising: (1) a Cas9 nickase; (2) an optional anchor DNA-binding domain (*e.g.*, ZFP) that specifically binds to a target site 5' or 3' to the Cas9 nickase; and (3) a non-Cas9 nickase (*e.g.*, a ZFP-nickase) comprising an A or C deaminase. The anchor DNA-binding domain and non-Cas9 nickase may bind on the same or different strands of the target and/or on the same (5' or 3') or different (one on 5' and one on 3') sides of the Cas9 nickase. In certain embodiments, the anchor DNA-binding domain and the non-Cas9 nickase base editor bind to opposite strands on opposite sides of the Cas9 nickase. *See, e.g.*, Figure 1B, bottom middle schematic.

[0115] In other embodiments, described herein is a base editing system comprising: (1) a Cas9 protein (*e.g.*, dCas9) operably linked to a sgRNA; (2) an optional anchor DNA-binding domain (*e.g.*, ZFP anchor) that specifically binds to a target site 5' or 3' to the Cas9 nickase; (3) a fusion protein comprising a ZFP operably linked to an A or C deaminase, which fusion protein is 3' or 5' to the Cas9 protein; and (4) a ZFN nickase that binds 3' or 5' of the Cas9 protein and/or the ZFP. The anchor DNA-binding domain and non-Cas9 protein may bind on the same or different strands of the target and/or on the same (5' or 3') or different (one on 5' and one on 3') sides of the Cas9 protein. In certain embodiments, the ZFP of the ZFP-deaminase fusion protein and the optional anchor ZFP bind to opposite strands. *See, e.g.*, Figure 1C. In further embodiments, the base editor does not comprise a ZFN nickase.

[0116] The Cas9 base editors described herein provide surprising and unexpected advantages in terms of PAM sequences that may be used for efficient and targeted base editing, including expanding (relaxing) the available PAM sequence for base editors comprising sgRNAs.

[0117] Also described herein are base editors (ABEs or CBEs) that do not comprise a Cas9 base editor (*e.g.*, lack a Cas9 nickase or Cas9 protein). *See, e.g.*, Figure 1A through Figure 1D.

[0118] In certain embodiments, the base editor comprises: (1) a non-Cas9 nickase, for example a ZFN nickase comprising a pair of ZFNs (ZFP operably linked to a nuclease domain) in which one of the nuclease domains of the pair is catalytically inactive (*see, e.g.*, U.S. Patent Nos. 8,703,489; 9,200,266; 9,631,186; and 10,113,207); and (2) a ZFP base editor comprising a ZFP operably linked to an A or C deaminase. *See, e.g.*, Figure 1B, bottom left schematic.

[0119] In other embodiments, the base editor comprises a DNA destabilizing molecule comprising any RNA-programmable molecule. In certain embodiments, the DNA destabilizing molecule comprises an RNA-programmable molecule comprising Cas9 protein (*e.g.*, dCas9) and sgRNA. In other embodiments, the RNA-programmable molecule is not a Cas9 protein (*e.g.*, Cpf1 (also known as Cas12a), C2c1, C2c2 (also known as Cas13a), C2c3, Cas1, Cas2, Cas4, CasX and CasY); and an adenosine or cytosine deaminase. Optionally, the base editor further comprises at least one ZFP DNA-binding domain (*e.g.*, any combination of a ZFP DNA-binding domain operably linked to the adenosine or cytosine deaminase; a ZFP anchor on either side of the DNA destabilizing molecule; and/or a ZFN nickase).

[0120] In other embodiments, the Cas9-free base editor comprises a ZFN nickase, and ABE or CBE (*e.g.* operably linked to a ZFP) and one or more DNA destabilizing molecules that makes the target base accessible (*e.g.*, unwinds the DNA). *See, e.g.*, Figure 1D. Non-limiting examples of DNA destabilizing (unwinding) molecules include protein domains as shown in Table A and nucleic acids, including LNAs and/or PNAs as shown in Figure 8. The ZFN nickase may include one or more mutations in the catalytically active and/or catalytically inactive *Fok*I domains and/or one or more mutations to the ZFP backbone. *See, e.g.*, U.S. Patent Publication No. 2018/0087072. In certain embodiments, the base editors as described herein may be Cas9-free but may include non-Cas9 CRISPR proteins.

[0121] Any of the Cas9-free (*e.g.*, ZFP) base editors may further comprise (or recruit) one or more additional DNA destabilizing factors (*e.g.*, DNA helicases, helix-destabilizing molecules and the bacterial DnaA protein, single-strand DNA binding proteins, oligonucleotides, etc.), for example if further unwinding of the DNA augments base editor function by increasing accessibility of the target. The one or

more DNA destabilizing factors can be associated with the nickase and/or the A or C deaminase-containing molecule. In certain embodiments, the DNA destabilizing factor (*e.g.*, DNA oligo) is associated with the ZFN nickase as depicted in Figure 1D.

[0122] In other embodiments, described herein is a base editing system comprising: (1) at least one DNA destabilizing molecule (*e.g.*, non-Cas9 protein); (2) an optional DNA-binding domain (*e.g.*, ZFN anchor) that specifically binds to a target site 5' or 3' to the DNA destabilizing molecule; and (3) a fusion protein comprising a ZFP operably linked to an A or C deaminase, which fusion protein is 3' or 5' to the DNA-destabilizing molecule.

[0123] Cas9-free base editors offer several surprising and unexpected advantages over Cas9 base editors including, but not limited to: (i) No Cas9 dependent off-target effects as we can build ZFNs with 99% or more cutting efficiency (no off-target effects); (ii) eliminating PAM restrictions as ZFPs can target essentially any sequence in the human genome; (iii) reducing and/or eliminating bystander mutations (*e.g.*, within the target window) prevalent with current Cas9 base editors; and/or (iv) facilitate AAV delivery (known to be functional *in vivo*) due to reduced construct size.

[0124] Thus, cas9-free base editors provide significant advantages over conventional Cas9 base editors including but not limited to: target specificity; versatility; control or elimination of bystander mutations; and ease of delivery. In terms of target specificity, Cas9-free ZFP base editors can be designed that are 99% efficient at editing with few or no off-target effects, whereas Cas9 base editors exhibit higher rates of off-target effects. Similarly, non-Cas9 ZFP base editors control (reduce) or eliminate bystander mutations as seen with Cas9 base editors. Furthermore, whereas the selection of target sites for non-Cas9 ZFP base editors is limited by PAM requirements, Cas9 free ZFP base editors can target essentially any target sequence). In addition, due to the reduced construct size of non-Cas9 base editors, they can be delivered using AAV vectors, thereby greatly expanding therapeutic (*in vivo*) uses.

[0125] In certain embodiments, the base editors (base editing systems) described herein include a single base to be changed (target base) within: (1) the base editing window and/or (2) bases between the base editing window and the 5' end of the PAM sequence intervening. *See,* also Figure 2. The opening of the DNA editing window leaves other non-target bases (adenines and/or cytosines) open for potential

modification by base editors. This means that although the target may be 13-16 nucleotides away from the 5' end of the PAM sequence, targeted bases (*e.g.*, adenines or cytosines) present in those intervening nucleotides and/or in the base editing window may also be altered. Such non-targeted mutations, also referred to as "bystander" mutations, which may be undesirable in the base editing process. In some situations, bystander mutations may be avoided by choosing a PAM sequence that do not comprise the base to be targeted (an adenine for adenine base editors or cytidine for cytidine base editor) in the editing window. Alternatively, and/or in addition to selection of a PAM sequences, the present invention reduces or eliminates bystander mutations by using a base editor as described herein comprising a ZFP anchor paired with a ZFN nickase functional domains, eliminating the PAM requirement, and allowing the user to place the editor at any optimal location.

[0126] In some embodiments, disclosed herein are complexes (systems) comprising a variety of fusion proteins. In some embodiments, Cas proteins can be fused to alternate DNA binding domains to increase the specificity of binding of the fusion protein to a DNA (see Bolukbashi, *et al.* (2015) *Nat Methods* 12(12):1150-1156). For example, ZFP, TtAgo and TALE DNA binding domains may be fused to a dCas. In some embodiments, the dCas comprises mutations to alter PAM specificity (see Gao, *et al.* (2017) *Nature Biotechnology* 35:789-792; Virginijus Siksnys (2016) *Mol Cell* 61:793) or to alter the requirement for PAM recognition. In some embodiments, the base editor lacks a Cas nuclease.

[0127] Potential targets for this approach are many. Non-limiting examples of base editing for treatment and/or prevention of disease editing of gene involved in exemplary diseases that may be treated include sickle cell disease, hemophilia, cystic fibrosis, phenylketonuria, Tay-Sachs, color blindness, Fabry disease, Friedreich's ataxia, prostate cancer, and many others.

[0128] Thus, the base editing systems as described herein may be used to alter expression of any disease-associate gene. In certain embodiments, the gene is associated with a cancer, for example, the JAK2 V617F mutation. This mutation plays a critical role in the expansion of myeloproliferative neoplasms. JAK2 transduces cytokine and growth factor signals from membrane-bound receptors through phosphorylation of the STAT family of transcription factors. The V617F mutation leads to constitutive tyrosine phosphorylation activity and promotes

cytokine hypersensitivity (James, *et al.* (2005) *Nature* 434:1144-1148) and the ability to drive cells to proliferate in the absence of cytokines (Zhao, *et al.* (2005) *J. Biol Chem* 280 (24):22788-22792). For some JAK2 V617F disorders (e.g. primary myelofibrosis), the only cure is hematopoietic cell transplant, however current approaches are often associated with disease relapse and graft versus host disease (Byrne, *et al.* (2018) *Ther Avd Hematol* 9(9):251-259). Editing of a subject's hematopoietic stem cells/progenitor cells (HSC/PC) to remove the mutation may allow successful treatment of these diseases.

[0129] In certain embodiments, base targets an alpha-1 antitrypsin (within the SERPINA locus). Mutations in the locus that cause an autosomal recessive deficiency in the A1AT protein are associated with both liver and lung disease. The PiZ mutation, one of the most common deficiency alleles in people of Northern Europe descent, results in only about 10-20% of the A1AT protein being produced. This mutation is caused by a single mutation in exon 5, leading to a glutamine substitution at amino acid position 342 for a lysine where a G at position 1096 in the DNA is an A in the mutated gene sequence (reviewed in Fregonese and Stolk (2008) *Orphanet J Rare Dis* 3:16).

DNA-binding molecules/domains

[0130] Described herein are compositions comprising one or more DNA-binding molecules/domains that specifically bind to a target site in any gene or locus of interest. Any DNA-binding molecules/domains can be used in the compositions and methods disclosed herein, including but not limited to a zinc finger DNA-binding domain, a TALE DNA binding domain, the DNA-binding portion (guide or sgRNA) of a CRISPR/Cas nuclease, and/or a DNA-binding domain from a meganuclease.

[0131] In certain embodiments, the base editors described herein comprise a zinc finger protein DNA-binding domain. Preferably, the zinc finger protein is non-naturally occurring in that it is engineered to bind to a target site of choice. *See*, for example, Beerli, *et al.* (2002) *Nature Biotechnol.* 20:135-141; Pabo, *et al.* (2001) *Ann. Rev. Biochem.* 70:313-340; Isalan, *et al.* (2001) *Nature Biotechnol.* 19:656-660; Segal, *et al.* (2001) *Curr. Opin. Biotechnol.* 12:632-637; Choo, *et al.* (2000) *Curr. Opin. Struct. Biol.* 10:411-416; U.S. Patent Nos. 6,453,242; 6,534,261; 6,599,692; 6,503,717; 6,689,558; 7,030,215; 6,794,136; 7,067,317; 7,262,054; 7,070,934; 7,361,635; and 7,253,273; and U.S. Patent Publication Nos. 2005/0064474;

2007/0218528; and 2005/0267061, all incorporated herein by reference in their entireties. In certain embodiments, the DNA-binding domain comprises a zinc finger protein disclosed in U.S. Patent Publication No. 2012/0060230, incorporated by reference in its entirety herein.

[0132] An engineered zinc finger binding domain can have a novel binding specificity, compared to a naturally-occurring zinc finger protein. Engineering methods include, but are not limited to, rational design and various types of selection. Rational design includes, for example, using databases comprising triplet (or quadruplet) nucleotide sequences and individual zinc finger amino acid sequences, in which each triplet or quadruplet nucleotide sequence is associated with one or more amino acid sequences of zinc fingers which bind the particular triplet or quadruplet sequence. See, for example, U.S. Patent Nos. 6,453,242 and 6,534,261, incorporated by reference herein in their entireties.

[0133] Exemplary selection methods, including phage display and two-hybrid systems, are disclosed in U.S. Patent Nos. 5,789,538; 5,925,523; 6,007,988; 6,013,453; 6,410,248; 6,140,466; 6,200,759; and 6,242,568; as well as International Patent Publication Nos. WO 98/37186; WO 98/53057; WO 00/27878; and WO 01/88197 and GB 2,338,237. In addition, enhancement of binding specificity for zinc finger binding domains has been described, for example, in U.S. Patent No. 6,794,136.

[0134] In addition, as disclosed in these and other references, zinc finger domains and/or multi-fingered zinc finger proteins may be linked together using any suitable linker sequences, including for example, linkers of 5 or more amino acids in length. See, also, U.S. Patent Nos. 6,479,626; 6,903,185; and 7,153,949 for exemplary linker sequences 6 or more amino acids in length. The proteins described herein may include any combination of suitable linkers between the individual zinc fingers of the protein. In addition, enhancement of binding specificity for zinc finger binding domains has been described, for example, in U.S. Patent No. 6,794,136.

[0135] Selection of target sites; ZFPs and methods for design and construction of fusion proteins (and polynucleotides encoding same) are known to those of skill in the art and described in detail in U.S. Patent Nos. 6,140,081; 5,789,538; 6,453,242; 6,534,261; 5,925,523; 6,007,988; 6,013,453; and 6,200,759; and International Patent Publication Nos. WO 95/19431; WO 96/06166; WO 98/53057; WO 98/54311;

WO 00/27878; WO 01/60970; WO 01/88197; WO 02/099084; WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536; and WO 03/016496.

[0136] In addition, as disclosed in these and other references, zinc finger domains and/or multi-fingered zinc finger proteins may be linked together using any suitable linker sequences, including for example, linkers of 5 or more amino acids in length. See, also, U.S. Patent Nos. 6,479,626; 6,903,185; and 7,153,949 for exemplary linker sequences 6 or more amino acids in length. The proteins described herein may include any combination of suitable linkers between the individual zinc fingers of the protein.

[0137] Usually, the ZFPs include at least three fingers. Certain of the ZFPs include four, five, six or more fingers. The ZFPs that include three fingers typically recognize a target site that includes 9 or 10 nucleotides; ZFPs that include four fingers typically recognize a target site that includes 12 to 14 nucleotides; while ZFPs having six fingers can recognize target sites that include 18 to 21 nucleotides. The ZFPs can also be fusion proteins that include one or more regulatory domains, which domains can be transcriptional activation or repression domains.

[0138] In some embodiments, the DNA-binding domain may be derived from a nuclease. For example, the recognition sequences of homing endonucleases and meganucleases such as I-*Sce*I, I-*Ceu*I, PI-*Psp*I, PI-*Sce*, I-*Sce*IV, I-*Csm*I, I-*Pan*I, I-*Sce*II, I-*Ppo*I, I-*Sce*III, I-*Cre*I, I-*Tev*I, I-*Tev*II and I-*Tev*III are known. *See* also U.S. Patent No. 5,420,032; U.S. Patent No. 6,833,252; Belfort, *et al.* (1997) *Nucleic Acids Res.* 25:3379-3388; Dujon, *et al.* (1989) *Gene* 82:115-118; Perler, *et al.* (1994) *Nucleic Acids Res.* 22, 1125-1127; Jasin (1996) *Trends Genet.* 12:224-228; Gimble, *et al.* (1996) *J. Mol. Biol.* 263:163-180; Argast, *et al.* (1998) *J. Mol. Biol.* 280:345-353 and the New England Biolabs catalogue. In addition, the DNA-binding specificity of homing endonucleases and meganucleases can be engineered to bind non-natural target sites. *See*, for example, Chevalier, *et al.* (2002) *Molec. Cell* 10:895-905; Epinat, *et al.* (2003) *Nucleic Acids Res.* 31:2952-2962; Ashworth, *et al.* (2006) *Nature* 441:656-659; Pâques, *et al.* (2007) *Current Gene Therapy* 7:49-66; U.S. Patent Publication No. 2007/0117128.

[0139] In certain embodiments, the zinc finger protein used with the mutant cleavage domains described herein comprises one or more mutations (substitutions, deletions, and/or insertions) to the backbone regions (*e.g.*, regions outside the 7-amino acid recognition helix region numbered -1 to 6), for example at one or more of

positions -14, -9 and/or -5. The wild-type residue at one or more these positions may be deleted, replaced with any amino acid residue and/or include on or more additional residues. In some embodiments, the Arg (R) at position -5 is changed to a Tyr (Y), Asp (N), Glu (E), Leu (L), Gln (Q), or Ala (A). In other embodiments, the Arg (R) at position (-9) is replaced with Ser (S), Asp (N), or Glu (E). In further embodiments, the Arg (R) at position (-14) is replaced with Ser (S) or Gln (Q). In other embodiments, the fusion polypeptides can comprise mutations in the zinc finger DNA binding domain where the amino acids at the (-5), (-9) and/or (-14) positions are changed to any of the above listed amino acids in any combination.

[0140] In other embodiments, the DNA binding domain comprises an engineered domain from a Transcriptional Activator-Like (TAL) effector (TALE) similar to those derived from the plant pathogens *Xanthomonas* (see Boch, *et al.* (2009) *Science* 326:1509-1512 and Moscou and Bogdanove (2009) *Science* 326:1501) and Ralstonia (see Heuer, *et al.* (2007) *Applied and Environmental Microbiology* 73(13):4379-4384); U.S. Patent Publication Nos. 2011/0301073 and 2011/0145940. The plant pathogenic bacteria of the genus *Xanthomonas* are known to cause many diseases in important crop plants. Pathogenicity of *Xanthomonas* depends on a conserved type III secretion (T3S) system which injects more than 25 different effector proteins into the plant cell. Among these injected proteins are transcription activator-like effectors (TALE) which mimic plant transcriptional activators and manipulate the plant transcriptome (see Kay, *et al.* (2007) *Science* 318:648-651). These proteins contain a DNA binding domain and a transcriptional activation domain. One of the most well characterized TALEs is AvrBs3 from *Xanthomonas campestgris* pv. *Vesicatoria* (see Bonas, *et al.* (1989) *Mol Gen Genet* 218:127-136 and International Patent Publication No. WO 2010/079430). TALEs contain a centralized domain of tandem repeats, each repeat containing approximately 34 amino acids, which are key to the DNA binding specificity of these proteins. In addition, they contain a nuclear localization sequence and an acidic transcriptional activation domain (for a review see Schornack S., *et al.* (2006) *J Plant Physiol* 163(3): 256-272). In addition, in the phytopathogenic bacteria *Ralstonia solanacearum* two genes, designated brg11 and hpx17 have been found that are homologous to the AvrBs3 family of *Xanthomonas* in the *R. solanacearum* biovar 1 strain GMI1000 and in the biovar 4 strain RS1000 (*See* Heuer, *et al.* (2007) *Appl and Envir Micro* 73(13):4379-4384). These genes are 98.9% identical in nucleotide sequence to each other but differ

by a deletion of 1,575 base pairs in the repeat domain of hpx17. However, both gene products have less than 40% sequence identity with AvrBs3 family proteins of *Xanthomonas*.

[0141] Specificity of these TAL effectors depends on the sequences found in the tandem repeats. The repeated sequence comprises approximately 102 base pairs and the repeats are typically 91-100% homologous with each other (Bonas, *et al.*, *ibid*). Polymorphism of the repeats is usually located at positions 12 and 13 and there appears to be a one-to-one correspondence between the identity of the hypervariable diresidues (the repeat variable diresidue or RVD region) at positions 12 and 13 with the identity of the contiguous nucleotides in the TAL-effector's target sequence (see Moscou and Bogdanove (2009) *Science* 326:1501 and Boch, *et al.* (2009) *Science* 326:1509-1512). Experimentally, the natural code for DNA recognition of these TAL-effectors has been determined such that an HD sequence at positions 12 and 13 (Repeat Variable Diresidue or RVD) leads to a binding to cytosine (C), NG binds to T, NI to A, C, G or T, NN binds to A or G, and ING binds to T. These DNA binding repeats have been assembled into proteins with new combinations and numbers of repeats, to make artificial transcription factors that are able to interact with new sequences and activate the expression of a non-endogenous reporter gene in plant cells (Boch, *et al.*, *ibid*). Engineered TAL proteins have been linked to a *Fok*I cleavage half domain to yield a TAL effector domain nuclease fusion (TALEN), including TALENs with atypical RVDs. *See, e.g.,* U.S. Patent No. 8,586,526.

[0142] In some embodiments, the TALEN comprises an endonuclease (*e.g.*, *Fok*I) cleavage domain or cleavage half-domain. In other embodiments, the TALE-nuclease is a mega TAL. These mega TAL nucleases are fusion proteins comprising a TALE DNA binding domain and a meganuclease cleavage domain. The meganuclease cleavage domain is active as a monomer and does not require dimerization for activity. (See Boissel *et al.* (2013) *Nucl Acid Res*: 1-13, doi: 10.1093/nar/gkt1224).

[0143] In still further embodiments, the nuclease comprises a compact TALEN. These are single chain fusion proteins linking a TALE DNA binding domain to a TevI nuclease domain. The fusion protein can act as either a nickase localized by the TALE region, or can create a double strand break, depending upon where the TALE DNA binding domain is located with respect to the TevI nuclease domain (see Beurdeley, *et al.* (2013) *Nat Comm*: 1-8 DOI:10.1038/ncomms2782). In

addition, the nuclease domain may also exhibit DNA-binding functionality. Any TALENs may be used in combination with additional TALENs (*e.g.*, one or more TALENs (cTALENs or *Fok*I-TALENs) with one or more mega-TALEs.

[0144] In addition, as disclosed in these and other references, zinc finger domains and/or multi-fingered zinc finger proteins or TALEs may be linked together using any suitable linker sequences, including for example, linkers of 5 or more amino acids in length. See, also, U.S. Patent Nos. 6,479,626; 6,903,185; and 7,153,949 for exemplary linker sequences 6 or more amino acids in length. The proteins described herein may include any combination of suitable linkers between the individual zinc fingers of the protein. In addition, enhancement of binding specificity for zinc finger binding domains has been described, for example, in U.S. Patent No. 6,794,136.

[0145] In certain embodiments, the base editor comprises a DNA-binding domain that is part of a CRISPR/Cas nuclease system, including a single guide RNA (sgRNA) DNA binding molecule that binds to DNA. *See, e.g.*, U.S. Patent No. 8,697,359 and U.S. Patent Publication Nos. 2015/0056705 and 2015/0159172. The CRISPR (clustered regularly interspaced short palindromic repeats) locus, which encodes RNA components of the system, and the cas (CRISPR-associated) locus, which encodes proteins (Jansen, *et al.* (2002) *Mol. Microbiol.* 43:1565-1575; Makarova, *et al.* (2002) *Nucleic Acids Res.* 30:482-496; Makarova, *et al.* (2006) *Biol. Direct* 1:7; Haft, *et al.* (2005) *PLoS Comput. Biol.* 1:e60) make up the gene sequences of the CRISPR/Cas nuclease system. CRISPR loci in microbial hosts contain a combination of CRISPR-associated (Cas) genes as well as non-coding RNA elements capable of programming the specificity of the CRISPR-mediated nucleic acid cleavage.

[0146] In some embodiments, the DNA binding domain is part of a TtAgo system (see Swarts, *et al.* (2014) *Nature* 507(7491):258-261; Swarts, *et al.* (2012) *PLoS* One 7(4):e35888; Sheng, *et al.*, *ibid*). In eukaryotes, gene silencing is mediated by the Argonaute (Ago) family of proteins. In this paradigm, Ago is bound to small (19-31 nt) RNAs. This protein-RNA silencing complex recognizes target RNAs via Watson-Crick base pairing between the small RNA and the target and endonucleolytically cleaves the target RNA (Vogel (2014) *Science* 344:972-973). In contrast, prokaryotic Ago proteins bind to small single-stranded DNA fragments and likely function to detect and remove foreign (often viral) DNA (Yuan, *et al.* (2005)

Mol. Cell 19:405; Olovnikov, *et al.* (2013) *Mol. Cell* 51, 594; Swarts, *et al.* (2014) *Nature* 507(7491):258-261; Swarts, *et al.* (2012) *PLoS* One 7(4):e35888). Exemplary prokaryotic Ago proteins include those from *Aquifex aeolicus*, *Rhodobacter sphaeroides*, and *Thermus thermophilus*.

[0147] One of the most well-characterized prokaryotic Ago protein is the one from *T. thermophilus* (TtAgo; Swarts, *et al.* (2014) *Nature* 507(7491):258-261; Swarts, *et al.* (2012) *PLoS* One 7(4):e35888). TtAgo associates with either 15 nt or 13-25 nt single-stranded DNA fragments with 5' phosphate groups. This "guide DNA" bound by TtAgo serves to direct the protein-DNA complex to bind a Watson-Crick complementary DNA sequence in a third-party molecule of DNA. Once the sequence information in these guide DNAs has allowed identification of the target DNA, the TtAgo-guide DNA complex cleaves the target DNA. Such a mechanism is also supported by the structure of the TtAgo-guide DNA complex while bound to its target DNA (G. Sheng, *et al., ibid*). Ago from *Rhodobacter sphaeroides* (RsAgo) has similar properties (Olovnikov, *et al., ibid*).

[0148] Exogenous guide DNAs of arbitrary DNA sequence can be loaded onto the TtAgo protein (Swarts, *et al.* (2014) *Nature* 507(7491):258-261; Swarts, *et al.* (2012) *PLoS* One 7(4):e35888). Since the specificity of TtAgo cleavage is directed by the guide DNA, a TtAgo-DNA complex formed with an exogenous, investigator-specified guide DNA will therefore direct TtAgo target DNA cleavage to a complementary investigator-specified target DNA. In this way, one may create a targeted double-strand break in DNA. Use of the TtAgo-guide DNA system (or orthologous Ago-guide DNA systems from other organisms) allows for targeted cleavage of genomic DNA within cells. Such cleavage can be either single- or double-stranded. For cleavage of mammalian genomic DNA, it would be preferable to use of a version of TtAgo codon optimized for expression in mammalian cells. Further, it might be preferable to treat cells with a TtAgo-DNA complex formed *in vitro* where the TtAgo protein is fused to a cell-penetrating peptide. Further, it might be preferable to use a version of the TtAgo protein that has been altered via mutagenesis to have improved activity at 37°C. Ago-RNA-mediated DNA cleavage could be used to affect a panoply of outcomes including gene knock-out, targeted gene addition, gene correction, targeted gene deletion using techniques standard in the art for exploitation of DNA breaks.

[0149] Thus, any DNA-binding molecule/domain can be used. In certain embodiments, the base editors described herein are Cas9 base editors that include a sgRNA DNA-binding domain (*e.g.*, as a part of a Cas9 nickase) and optionally, one or more ZFP DNA-binding domains (referred to as "ZFP anchors"), which ZFP(s) can increase base editing efficiency and/or specificity. Non-limiting examples of Cas9 base editors including ZFP anchors are shown in Figure 1B and Figure 3.

Fusion molecules

[0150] The DNA-editing complexes described herein can include one or more fusion molecules comprising DNA-binding domains (*e.g.*, ZFPs or TALEs, CRISPR/Cas components such as single guide RNAs) as described herein and a heterologous (functional) domain (or functional fragment thereof) are also provided.

[0151] Common domains include, *e.g.*, transcription factor domains (activators, repressors, co-activators, co-repressors), silencers, oncogenes (*e.g.*, myc, jun, fos, myb, max, mad, rel, ets, bcl, myb, mos family members etc.); DNA repair enzymes and their associated factors and modifiers; helicases, double strand DNA binding proteins, DNA rearrangement enzymes and their associated factors and modifiers; chromatin associated proteins and their modifiers (*e.g.* kinases, acetylases and deacetylases); and DNA modifying enzymes (*e.g.*, methyltransferases, topoisomerases, helicases, ligases, kinases, phosphatases, deaminases, polymerases, endonucleases) and their associated factors and modifiers. U.S. Patent Publication Nos. 2005/0064474; 2006/0188987; and 2007/0218528 for details regarding fusions of DNA-binding domains and nuclease cleavage domains, incorporated by reference in their entireties herein.

[0152] Fusion molecules are constructed by methods of cloning and biochemical conjugation that are well known to those of skill in the art. Fusion molecules comprise a DNA-binding domain and a functional domain (for example, a helicase and/or deaminase and/or a GUI and/or GAM). Fusion molecules also optionally comprise nuclear localization signals (such as, for example, that from the SV40 medium T-antigen) and epitope tags (such as, for example, FLAG and hemagglutinin). Fusion proteins (and nucleic acids encoding them) are designed such that the translational reading frame is preserved among the components of the fusion.

[0153] Fusions between a polypeptide component of a functional domain (or a functional fragment thereof) on the one hand, and a non-protein DNA-binding domain

(e.g., antibiotic, intercalator, minor groove binder, nucleic acid) on the other, are constructed by methods of biochemical conjugation known to those of skill in the art. See, for example, the Pierce Chemical Company (Rockford, IL) Catalogue. Methods and compositions for making fusions between a minor groove binder and a polypeptide have been described. Mapp, *et al.* (2000) *Proc. Natl. Acad. Sci. USA* 97:3930-3935. Furthermore, single guide RNAs of the CRISPR/Cas system associate with functional domains to form active transcriptional regulators and nucleases.

[0154] In certain embodiments, the target site for the DNA-binding domain is present in an accessible region of cellular chromatin. Accessible regions can be determined as described, for example, in U.S. Patent Nos. 7,217,509 and 7,923,542. If the target site is not present in an accessible region of cellular chromatin, one or more accessible regions can be generated as described in U.S. Patent Nos. 7,785,792 and 8,071,370. In additional embodiments, the DNA-binding domain of a fusion molecule is capable of binding to cellular chromatin regardless of whether its target site is in an accessible region or not. For example, such DNA-binding domains are capable of binding to linker DNA and/or nucleosomal DNA. Examples of this type of "pioneer" DNA binding domain are found in certain steroid receptor and in hepatocyte nuclear factor 3 (HNF3) (Cordingley, *et al.* (1987) *Cell* 48:261-270; Pina, *et al.* (1990) *Cell* 60:719-731; and Cirillo, *et al.* (1998) *EMBO J.* 17:244-254).

[0155] The fusion molecule may be formulated with a pharmaceutically acceptable carrier, as is known to those of skill in the art. See, for example, Remington's Pharmaceutical Sciences, 17th ed., 1985; and U.S. Patent Nos. 6,453,242 and 6,534,261.

[0156] The functional component(s)/domain(s) of a fusion molecule can be selected from any of a variety of different components capable of influencing the sequence of a gene once the fusion molecule binds to a target sequence via its DNA binding domain. Hence, the functional component can include, but is not limited to, various deaminases, UGI, GAM, helicases etc. In certain embodiments, the functional domain comprises one or more cytidine deaminases (*e.g.*, an apolipoprotein B mRNA-editing complex 1 (APOBEC1) domain and/or an Activation Induced Deaminase (AID)). In other embodiments, the functional domain comprises one or adenine deaminases (*e.g.*, a mutated TadA (tRNA adenine deaminase (*see* Gaudelli, *et al.* (2017) *Nature* 551:464-471)). In still further embodiments, the functional domain

comprises at least one uracil DNA glycosylase inhibitor (*e.g.* UGI) domain. In some embodiments, the base editing complex comprises a deaminase, a nickase, a UGI and/or a GAM protein. The functional domain(s) may be positioned with respect to the DNA-binding domain (and/or nickase when included in the catalytically active fusion molecule) in any way including but not limited to N-terminal (in any order when multiple functional domains are present), C-terminal (in any order when multiple functional domains are present), etc.

[0157] In some embodiments, the DNA-editing (base editing) complex further comprises a molecule to assist in opening or destabilizing a double strand DNA helix. In some embodiments, the molecule comprises an enzyme. In some embodiments, the enzyme is a helicase (for example, RecQ helicases (WRN, BLM, RecQL4 and RecQ5, (see Mo, *et al.* (2018) *Cancer Lett.* 413:1-10), DNA2 (Jia, *et al.* (2017) *DNA Repair* (Amst). 59:9-19) and any other eukaryotic helicases including for example, FANCJ, XPD, XPB, RTEL1, and PIF1 (Brosh (2013) *Nat Rev Canc* 13(8):542-558)). In some embodiments, the enzyme is a bacterial and/or a viral helicase. Exemplary viral helicases include those encoded by the Myoviridae family of viruses (for example gp41, Dda, UvsW, Gene α, and Ban); those encoded by the Podpviridae family of viruses (for example 4B); those encoded by the Siphoviridae, Baculoviridae, Herpesviridae, Polyomaviridae, Palillomaviridae and Poxviridae families (for example, G40P, p143, UL5, UL9, Tag, E1, NPH-I, NPH-II, A18R, and VETF), or any other viral helicase known in the art (see e.g. Frick and Lam (2006) *Curr Pharm Des* 12(11):1315-1338). In some embodiments, the helicase enzyme is a bacterial enzyme. Exemplary bacterial helicases include the P. aeruginosa SF4 DnaB-like helicase, or the RecB and RecD helicases that are part of the bacterial RecBCD complex in bacteria such as E coli and H. pylori (Shadrick, *et al.* (2013) *J. Biomol Screen* 18(7):761-781). In some embodiments, the molecule comprises a CRISPR/Cas complex. In some embodiments, the CRISPR/Cas complex comprises a guide RNA. In some embodiments, the complex comprises a Cas enzyme that is catalytically defective in one of the nuclease domains. In some embodiments, the Cas enzyme is defective in its PAM recognition (Anders, *et al.* (2014) *Nature* 513(7519):569-573). In some embodiments, the molecule has helix-destabilizing properties. Exemplary helix-destabilizing molecules include ICP8 from herpes simplex virus type I (Boehmer and Lehman (1993) *J Virol* 67(2):711-715), Puralpha (Darbinian, *et al.* (2001) *J Cell Biochem* 80(4):589-95), and calf thymus DNA helix-destabilizing

protein (Kohwi-Shigematsu, *et al.* (1978) *Proc Natl Acad Sci USA* 75(10):4689-93). In some embodiments, the molecule is a nucleic acid. In some embodiments, the nucleic acid is a DNA with homology to the region near the targeted editing. In some embodiments, the nucleic acid is an RNA with homology to the region near the targeted editing. In some embodiments, the RNA is modified. In some embodiments, the fusion molecule comprises amino acid linker sequences between one or more domains of the fusion molecule.

[0158] The DNA-editing complexes described herein can include 1, 2, 3, 4 or more fusion molecules as described herein. In certain embodiments, the DNA-editing complex comprises 2 fusion molecules: a first fusion molecule that is a catalytically active nickase (catalytically active) comprising a DNA-binding domain and nickase domain and a second catalytically inactive fusion molecule comprising a DNA-binding domain and one or more functional domains (cytidine deaminase, adenine deaminase, and/or UGI, etc.). Typically, the fusion molecules are "partners" in that the two DNA-binding domains bind to target sites such that the two fusion molecules dimerize to effect DNA editing. In other embodiments, the DNA-editing complex comprises 3 or more fusion molecules: a first fusion molecule that is a catalytically active nickase comprising a DNA-binding domain and a nickase domain; a second catalytically inactive fusion molecule comprising a DNA-binding domain (*e.g.*, that is a partner and dimerizes with the first fusion molecule); and a third fusion molecule comprising a DNA-binding domain and one or more functional domains as described herein.

Nickase Domains

[0159] In certain embodiments, the fusion protein comprises a DNA-binding binding domain and cleavage (nuclease) domain, preferably a nickase domain. As such, gene editing can be achieved using a nuclease, for example an engineered nickase. Engineered nuclease technology is based on the engineering of naturally occurring DNA-binding proteins. For example, engineering of homing endonucleases with tailored DNA-binding specificities has been described. Chames, *et al.* (2005) *Nucleic Acids Res* 33(20):e178; Arnould, *et al.* (2006) *J. Mol. Biol.* 355:443-458. In addition, engineering of ZFPs has also been described. See, e.g., U.S. Patent Nos. 6,534,261; 6,607,882; 6,824,978; 6,979,539; 6,933,113; 7,163,824; and 7,013,219.

[0160] In addition, ZFPs and/or TALEs have been fused to nuclease domains to create ZFNs and TALENs – a functional entity that is able to recognize its intended nucleic acid target through its engineered (ZFP or TALE) DNA binding domain and cause the DNA to be cut near the DNA binding site via the nuclease activity. See, e.g., Kim, *et al.* (1996) *Proc Nat'l Acad Sci USA* 93(3):1156-1160. More recently, such nucleases have been used for genome modification in a variety of organisms. See, *e.g.*, U.S. Patent Nos. 9,255,250; 9,200,266; 9,045,763; 9,005,973; 8,956,828; 8,945,868; 8,703,489; 8,586,526; 6,534,261; 6,599,692; 6,503,717; 6,689,558; 7,067,317; 7,262,054; 7,888,121; 7,972,854; 7,914,796; 7,951,925; 8,110,379; 8,409,861; U.S. Patent Publication Nos. 2003/0232410; 2005/0208489; 2005/0026157; 2005/0064474; 2006/0063231; 2008/0159996; 2010/00218264; 2012/0017290; 2011/0265198; 2013/0137104; 2013/0122591; 2013/0177983; 2013/0177960; and 2015/0056705.

[0161] Thus, the methods and compositions described herein are broadly applicable and may involve any nuclease of interest. Non-limiting examples of nucleases include meganucleases, TALENs and zinc finger nucleases. The nuclease may comprise heterologous DNA-binding and cleavage domains (e.g., zinc finger nucleases; meganuclease DNA-binding domains with heterologous cleavage domains) or, alternatively, the DNA-binding domain of a naturally-occurring nuclease may be altered to bind to a selected target site (e.g., a meganuclease that has been engineered to bind to site different than the cognate binding site).

[0162] In any of the nucleases described herein, the nuclease can comprise an engineered TALE DNA-binding domain and a nuclease domain (e.g., endonuclease and/or meganuclease domain), also referred to as TALENs. Methods and compositions for engineering these TALEN proteins for robust, site specific interaction with the target sequence of the user`s choosing have been published (see U.S. Patent No. 8,586,526). In some embodiments, the TALEN comprises an endonuclease (e.g., *Fok*I) cleavage domain or cleavage half-domain. In other embodiments, the TALE-nuclease is a mega TAL. These mega TAL nucleases are fusion proteins comprising a TALE DNA binding domain and a meganuclease cleavage domain. The meganuclease cleavage domain is active as a monomer and does not require dimerization for activity. (See Boissel, *et al.* (2013) *Nucl Acid Res* 1-13, doi:10.1093/nar/gkt1224). In addition, the nuclease domain may also exhibit DNA-binding functionality.

[0163] In still further embodiments, the nuclease comprises a compact TALEN (cTALEN). These are single chain fusion proteins linking a TALE DNA binding domain to a TevI nuclease domain. The fusion protein can act as either a nickase localized by the TALE region, or can create a double strand break, depending upon where the TALE DNA binding domain is located with respect to the TevI nuclease domain (see Beurdeley, *et al.* (2013) *Nat Comm:* 1-8 DOI:10.1038/ncomms2782). Any TALENs may be used in combination with additional TALENs (e.g., one or more TALENs (cTALENs or *Fok*I-TALENs) with one or more mega-TALs) or other DNA cleavage enzymes.

[0164] In certain embodiments, the nuclease comprises a meganuclease (homing endonuclease) or a portion thereof that exhibits cleavage activity. Naturally-occurring meganucleases recognize 15-40 base-pair cleavage sites and are commonly grouped into four families: the LAGLIDADG family ("LAGLIDADG" disclosed as SEQ ID NO:75), the GIY-YIG family, the His-Cyst box family and the HNH family. Exemplary homing endonucleases include I-SceI, I-CeuI, PI-PspI, PI-Sce, I-SceIV, I-CsmI, I-PanI, I-SceII, I-PpoI, I-SceIII, I-CreI, I-TevI, I-TevII and I-TevIII. Their recognition sequences are known. See also U.S. Patent No. 5,420,032; U.S. Patent No. 6,833,252; Belfort, *et al.* (1997) *Nucleic Acids Res.* 25:3379-3388; Dujon, *et al.* (1989) *Gene* 82:115-118; Perler, *et al.* (1994) *Nucleic Acids Res.* 22:1125-1127; Jasin (1996) *Trends Genet.* 12:224-228; Gimble, *et al.* (1996) *J. Mol. Biol.* 263:163-180; Argast, *et al.* (1998) *J. Mol. Biol.* 280:345-353 and the New England Biolabs catalogue.

[0165] DNA-binding domains from naturally-occurring meganucleases, primarily from the LAGLIDADG family ("LAGLIDADG" disclosed as SEQ ID NO:75), have been used to promote site-specific genome modification in plants, yeast, Drosophila, mammalian cells and mice, but this approach has been limited to the modification of either homologous genes that conserve the meganuclease recognition sequence (Monet, *et al.* (1999) *Biochem. Biophysics. Res. Common* 255:88-93) or to pre-engineered genomes into which a recognition sequence has been introduced (Route, *et al.* (1994) *Mol. Cell. Biol.* 14:8096-106; Chilton, *et al.* (2003) *Plant Physiology* 133:956-65; Puchta, *et al.* (1996) *Proc. Natl. Acad. Sci. USA* 93:5055-60; Rong, *et al.* (2002) *Genes Dev.* 16:1568-81; Gouble, *et al.* (2006) *J. Gene Med.* 8(5):616-622). Accordingly, attempts have been made to engineer meganucleases to exhibit novel binding specificity at medically or biotechnologically relevant sites

(Porteus, *et al.* (2005) *Nat. Biotechnol.* 23:967-73; Sussman, *et al.* (2004) *J. Mol. Biol.* 342:31-41; Epinat, *et al.* (2003) *Nucleic Acids Res.* 31:2952-62; Chevalier, *et al.* (2002) *Molec. Cell* 10:895-905; Epinat, *et al.* (2003) *Nucleic Acids Res.* 31:2952-2962; Ashworth, *et al.* (2006) *Nature* 441:656-659; Paques, *et al.* (2007) *Current Gene Therapy* 7:49-66; U.S. Patent Publication Nos. 2007/0117128; 2006/0206949; 2006/0153826; 2006/0078552; and 2004/0002092). In addition, naturally-occurring or engineered DNA-binding domains from meganucleases can be operably linked with a cleavage domain from a heterologous nuclease (e.g., *Fok*I) and/or cleavage domains from meganucleases can be operably linked with a heterologous DNA-binding domain (e.g., ZFP or TALE).

[0166] In other embodiments, the nuclease is a zinc finger nuclease (ZFN) or TALE DNA binding domain-nuclease fusion (TALEN). ZFNs and TALENs comprise a DNA binding domain (zinc finger protein or TALE DNA binding domain) that has been engineered to bind to a target site in a gene of choice and cleavage domain or a cleavage half-domain (e.g., from a restriction and/or meganuclease as described herein).

[0167] As described in detail above, zinc finger binding domains and TALE DNA binding domains can be engineered to bind to a sequence of choice. See, for example, Beerli, *et al.* (2002) *Nature Biotechnol.* 20:135-141; Pabo, *et al.* (2001) *Ann. Rev. Biochem.* 70:313-340; Isalan, *et al.* (2001) *Nature Biotechnol.* 19:656-660; Segal, *et al.* (2001) *Curr. Opin. Biotechnol.* 12:632-637; Choo, *et al.* (2000) *Curr. Opin. Struct. Biol.* 10:411-416. An engineered zinc finger binding domain or TALE protein can have a novel binding specificity, compared to a naturally-occurring protein. Engineering methods include, but are not limited to, rational design and various types of selection. Rational design includes, for example, using databases comprising triplet (or quadruplet) nucleotide sequences and individual zinc finger or TALE amino acid sequences, in which each triplet or quadruplet nucleotide sequence is associated with one or more amino acid sequences of zinc fingers or TALE repeat units which bind the particular triplet or quadruplet sequence. See, for example, U.S. Patent Nos. 6,453,242 and 6,534,261, incorporated by reference herein in their entireties.

[0168] Selection of target sites; and methods for design and construction of fusion proteins (and polynucleotides encoding same) are known to those of skill in the

art and described in detail in U.S. Patent Nos. 7,888,121 and 8,409,861, incorporated by reference in their entireties herein.

[0169] In addition, as disclosed in these and other references, zinc finger domains, TALEs and/or multi-fingered zinc finger proteins may be linked together using any suitable linker sequences, including for example, linkers of 5 or more amino acids in length. See, e.g., U.S. Patent Nos. 6,479,626; 6,903,185; and 7,153,949 for exemplary linker sequences 6 or more amino acids in length. The proteins described herein may include any combination of suitable linkers between the individual zinc fingers of the protein. See, also, U.S. Patent No. 8,772,453.

[0170] Thus, nucleases such as ZFNs, TALENs and/or meganucleases can comprise any DNA-binding domain and any nuclease (cleavage) domain (cleavage domain, cleavage half-domain). As noted above, the cleavage domain may be heterologous to the DNA-binding domain, for example a zinc finger or TAL-effector DNA-binding domain and a cleavage domain from a nuclease or a meganuclease DNA-binding domain and cleavage domain from a different nuclease. Heterologous cleavage domains can be obtained from any endonuclease or exonuclease. Exemplary endonucleases from which a cleavage domain can be derived include, but are not limited to, restriction endonucleases and homing endonucleases. See, for example, 2002-2003 Catalogue, New England Biolabs, Beverly, MA; and Belfort, *et al.* (1997) *Nucleic Acids Res.* 25:3379-3388. Additional enzymes which cleave DNA are known (e.g., S1 Nuclease; mung bean nuclease; pancreatic DNase I; micrococcal nuclease; yeast HO endonuclease; see also Linn, *et al.* (eds.) Nucleases, Cold Spring Harbor Laboratory Press, 1993). One or more of these enzymes (or functional fragments thereof) can be used as a source of cleavage domains and cleavage half-domains.

[0171] Similarly, a cleavage half-domain can be derived from any nuclease or portion thereof, as set forth above, that requires dimerization for cleavage activity. In general, two fusion proteins are required for cleavage if the fusion proteins comprise cleavage half-domains. Alternatively, a single protein comprising two cleavage half-domains can be used. The two cleavage half-domains can be derived from the same endonuclease (or functional fragments thereof), or each cleavage half-domain can be derived from a different endonuclease (or functional fragments thereof). In addition, the target sites for the two fusion proteins are preferably disposed, with respect to each other, such that binding of the two fusion proteins to their respective target sites places the cleavage half-domains in a spatial orientation to each other that allows the

cleavage half-domains to form a functional cleavage domain, *e.g.*, by dimerizing. Thus, in certain embodiments, the near edges of the target sites are separated by 5-10 nucleotides or by 15-18 nucleotides. However, any integral number of nucleotides or nucleotide pairs can intervene between two target sites (e.g., from 2 to 50 nucleotide pairs or more). In general, the site of cleavage lies between the target sites.

[0172] Restriction endonucleases (restriction enzymes) are present in many species and are capable of sequence-specific binding to DNA (at a recognition site), and cleaving DNA at or near the site of binding. Certain restriction enzymes (e.g., Type IIS) cleave DNA at sites removed from the recognition site and have separable binding and cleavage domains. For example, the Type IIS enzyme *Fok*I catalyzes double-stranded cleavage of DNA, at 9 nucleotides from its recognition site on one strand and 13 nucleotides from its recognition site on the other. See, for example, U.S. Patent Nos. 5,356,802; 5,436,150; and 5,487,994; as well as Li, *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:4275-4279; Li, *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:2764-2768; Kim, *et al.* (1994a) *Proc. Natl. Acad. Sci. USA* 91:883-887; Kim, *et al.* (1994b) *J. Biol. Chem.* 269:31,978-31,982. Thus, in one embodiment, fusion proteins comprise the cleavage domain (or cleavage half-domain) from at least one Type IIS restriction enzyme and one or more zinc finger binding domains, which may or may not be engineered.

[0173] An exemplary Type IIS restriction enzyme, whose cleavage domain is separable from the binding domain, is *Fok*I. This particular enzyme is active as a dimer. Bitinaite, *et al.* (1998) *Proc. Natl. Acad. Sci. USA* 95:10,570-10,575. Accordingly, for the purposes of the present disclosure, the portion of the *Fok*I enzyme used in the disclosed fusion proteins is considered a cleavage half-domain. Thus, for targeted double-stranded cleavage and/or targeted replacement of cellular sequences using zinc finger-*Fok*I fusions, two fusion proteins, each comprising a *Fok*I cleavage half-domain, can be used to reconstitute a catalytically active cleavage domain. Alternatively, a single polypeptide molecule containing a zinc finger binding domain and two *Fok*I cleavage half-domains can also be used. Parameters for targeted cleavage and targeted sequence alteration using zinc finger-*Fok*I fusions are provided elsewhere in this disclosure.

[0174] A cleavage domain or cleavage half-domain can be any portion of a protein that retains cleavage activity, or that retains the ability to multimerize (e.g., dimerize) to form a functional cleavage domain.

[0175] Exemplary Type IIS restriction enzymes are described in International Patent Publication No. WO 07/014275, incorporated herein in its entirety. Additional restriction enzymes also contain separable binding and cleavage domains, and these are contemplated by the present disclosure. See, for example, Roberts, *et al.* (2003) *Nucleic Acids Res.* 31:418-420.

[0176] In certain embodiments, the cleavage domain is a *Fok*I cleavage domain. The full-length *Fok*I sequence is shown below. The cleavage domain is shown in italics and underlining (positions 384 to 579 of the full- length protein) where the holo protein sequence is described below (SEQ ID NO:5):

MVSKIRTFGWVQNPGKFENLKRVVQVFDRNSKVHNEVKNIKIPTLVKESKIQ
KELVAIMNQHDLIYTYKELVGTGTSIRSEAPCDAIIQATIADQGNKKGYIDNW
SSDGFLRWAHALGFIEYINKSDSFVITDVGLAYSKSADGSAIEKEILIEAISSYPP
AIRILTLLEDGQHLTKFDLGKNLGFSGESGFTSLPEGILLDTLANAMPKDKGEI
RNNWEGSSDKYARMIGGWLDKLGLVKQGKKEFIIPTLGKPDNKEFISHAFKIT
GEGLKVLRRAKGSTKFTRVPKRVYWEMLATNLTDKEYVRTRRALILEILIKA
GSLKIEQIQDNLKKLGFDEVIETIENDIKGLINTGIFIEIKGRFYQLKDHILQFVIP
NRGVTK*QLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFM*
KVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYV
EENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGA
VLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF

[0177] Cleavage half domains derived from *Fok*I may comprise a mutation in one or more of amino acid residues as shown in SEQ ID NO:5. Mutations include substitutions (of a wild-type amino acid residue for a different residue, insertions (of one or more amino acid residues) and/or deletions (of one or more amino acid residues). In certain embodiments, one or more of residues 414-426, 443-450, 467-488, 501-502, and/or 521-531 (numbered relative to SEQ ID NO:5) are mutated since these residues are located close to the DNA backbone in a molecular model of a ZFN bound to its target site described in Miller, *et al.* (2007) *Nat Biotechnol* 25:778-784). In certain embodiments, one or more residues at positions 416, 422, 447, 448, and/or 525 are mutated. In certain embodiments, the mutation comprises a substitution of a wild-type residue with any different residue, for example an alanine (A) residue, a cysteine (C) residue, an aspartic acid (D) residue, a glutamic acid (E) residue, a histidine (H) residue, a phenylalanine (F) residue, a glycine (G) residue, an asparagine (N) residue, a serine (S) residue or a threonine (T) residue. In other embodiments, the wild-type residue at one or more of positions 416, 418, 422, 446, 448, 476, 479, 480, 481, and/or 525 are replaced with any other residues, including but not limited to,

R416D, R416E, S418E, S418D, R422H, S446D, K448A, N476D, I479Q, I479T, G480D, Q481A, Q481E, K525S, K525A, N527D, R416E+R422H, R416D+R422H, R416E+K448A, R416D+R422H, K448A+I479Q, K448A+Q481A. K448A+K525A.

[0178] In certain embodiments, the cleavage domain comprises one or more engineered cleavage half-domain (also referred to as dimerization domain mutants) that minimize or prevent homodimerization, as described, for example, in U.S. Patent Nos. 7,914,796; 8,034,598; and 8,623,618; and U.S. Patent Publication No. 2011/0201055, the disclosures of all of which are incorporated by reference in their entireties herein. Amino acid residues at positions 446, 447, 479, 483, 484, 486, 487, 490, 491, 496, 498, 499, 500, 531, 534, 537, and 538 of *Fok*I (numbered relative to SEQ ID NO:5) are all targets for influencing dimerization of the *Fok*I cleavage half-domains. The mutations may include mutations to residues found in natural restriction enzymes homologous to *Fok*I. In a preferred embodiment, the mutation at positions 416, 422, 447, 448 and/or 525 (numbered relative to SEQ ID NO:5) comprise replacement of a positively charged amino acid with an uncharged or a negatively charged amino acid. In another embodiment, the engineered cleavage half domain comprises mutations in amino acid residues 499, 496 and 486 in addition to the mutations in one or more amino acid residues 416, 422, 447, 448, or 525, all numbered relative to SEQ ID NO:5.

[0179] Any nickase domain can be used in the DNA-editing complexes described herein. Nickases comprise mutations in a catalytic domain to render a nuclease unable to make a full double strand break, but instead result in the partial cleavage, "nicking" of a double stranded DNA. In embodiments in which two or more cleavage domains are necessary to nick the target, typically at least one of the cleavage domains (e.g., cleavage half-domains) includes one more mutations to its catalytic domain, which renders the nuclease inactive (e.g., catalytically inactive half domain). Catalytically inactive cleavage domains for producing nickases include but not are not limited to mutated *Fok*I and/or dCas proteins. See, e.g., U.S. Patent Nos. 9,522,936; 9,631,186; 9,200,266; and 8,703,489 and Guillinger, *et al.* (2014) *Nature Biotech.* 32(6):577-582; Cho, *et al.* (2014) *Genome Res.* 24(1):132-141). These catalytically inactive cleavage domain may, in combination with a catalytically active domain act as a nickase to make a single-stranded cut. Additional nickases are also known in the art, for example, McCaffery, *et al.* (2016) *Nucleic Acids Res.* 44(2):e11. doi:10.1093/nar/gkv878. Epub 2015 Oct 19.

[0180] In certain embodiments, the nickase comprises a nuclease nickase that comprises a catalytically inactive *Fok*I cleavage domain, for example a zinc finger nuclease (ZFN) nickase or a TAL-effector domain (TALEN) nickase. Non-limiting examples of amino acids that can be mutated in the catalytic domain of *Fok*I includes amino acid residues 450, 467 and/or 469 (as determined relative to wild-type). In certain embodiments, one or more point mutations are made in the catalytic domain of one member of the obligate heterodimer so as to inactivate the catalytic activity of the cleavage half-domain. For instance, position 450 may be mutated from D to N, position 467 may be mutated from D to A; and position 469 may be mutated from K to A. Other amino acids may be substituted at these or other positions. See, e.g., U.S. Patent Nos. 9,522,936; 9,631,186; 8,703,489 and 9,200,266 and Guillinger, *et al.* (2014) *Nature Biotech.* 32(6):577-582; Cho, *et al.* (2014) *Genome Res.* 24(1):132-141). The catalytically inactive cleavage domain may, in combination with a catalytically active domain act as a nickase to make a single-stranded cut. Additional nickases are also known in the art, for example, McCaffery, *et al.* (2016) *Nucleic Acids Res.* 44(2):e11. doi:10.1093/nar/gkv878. Epub 2015 Oct 19. Any nuclease (e.g., ZFN or TALEN or CRISPR/Cas nuclease) can become a nickase by using cleavage domains that make a single-stranded cut in place of the cleavage domains in nucleases that make double stranded cuts.

[0181] *Fok*I domains may also include one or more additional mutations. In certain embodiments, the compositions described herein include engineered cleavage half-domains of *Fok*I that form obligate heterodimers as described, for example, in U.S. Patent Nos. 7,914,796; 8,034,598; 8,961,281; and 8,623,618; U.S. Patent Publication Nos. 2008/0131962 and 2012/0040398. Thus, in one preferred embodiment, the invention provides fusion proteins wherein the engineered cleavage half-domain comprises a polypeptide in which the wild-type Gln (Q) residue at position 486 is replaced with a Glu (E) residue, the wild-type Ile (I) residue at position 499 is replaced with a Leu (L) residue and the wild-type Asn (N) residue at position 496 is replaced with an Asp (D) or a Glu (E) residue ("ELD" or "ELE") in addition to one or more mutations at positions 416, 422, 447, 448, or 525 (numbered relative to SEQ ID NO:5). In another embodiment, the engineered cleavage half domains are derived from a wild-type *Fok*I cleavage half domain and comprise mutations in the amino acid residues 490, 538 and 537, numbered relative to wild-type *Fok*I (SEQ ID NO:5) in addition to the one or more mutations at amino acid residues 416, 422, 447,

448, or 525. In a preferred embodiment, the invention provides a fusion protein, wherein the engineered cleavage half-domain comprises a polypeptide in which the wild-type Glu (E) residue at position 490 is replaced with a Lys (K) residue, the wild-type Ile (I) residue at position 538 is replaced with a Lys (K) residue, and the wild-type His (H) residue at position 537 is replaced with a Lys (K) residue or an Arg (R) residue ("KKK" or "KKR") (see U.S. 8,962,281, incorporated by reference herein) in addition to one or more mutations at positions 416, 422, 447, 448, or 525. See, e.g., U.S. Patent Nos. 7,914,796; 8,034,598; and 8,623,618, the disclosures of which are incorporated by reference in its entirety for all purposes. In other embodiments, the engineered cleavage half domain comprises the "Sharkey" and/or "Sharkey mutations" (see Guo, *et al.* (2010) *J. Mol. Biol.* 400(1):96-107).

[0182] In other embodiments, the nickases as described herein comprise engineered cleavage half domains are derived from a wild-type *Fok*I cleavage half domain and comprise mutations in the amino acid residues 490, and 538, numbered relative to wild-type *Fok*I or a *Fok*I homologue in addition to the one or more mutations at amino acid residues 416, 422, 447, 448, or 525. In preferred embodiments, the invention provides a fusion protein, wherein the engineered cleavage half-domain comprises a polypeptide in which the wild-type Glu (E) residue at position 490 is replaced with a Lys (K) residue, and the wild-type Ile (I) residue at position 538 is replaced with a Lys (K) residue ("KK") in addition to one or more mutations at positions 416, 422, 447, 448, or 525. In other preferred embodiments, the description provides a fusion protein, wherein the engineered cleavage half-domain comprises a polypeptide in which the wild-type Gln (Q) residue at position 486 is replaced with an Glu (E) residue, and the wild-type Ile (I) residue at position 499 is replaced with a Leu (L) residue ("EL") (See U.S. Patent No. 8,034,598, incorporated by reference herein) in addition to one or more mutations at positions 416, 422, 447, 448, or 525.

[0183] In some aspects, the description provides a fusion protein wherein the engineered cleavage half-domain comprises a polypeptide in which the wild-type amino acid residue at one or more of positions 387, 393, 394, 398, 400, 402, 416, 422, 427, 434, 439, 441, 447, 448, 469, 487, 495, 497, 506, 516, 525, 529, 534, 559, 569, 570, 571 in the *Fok*I catalytic domain are mutated. In some embodiments, the one or more mutations alter the wild type amino acid from a positively charged residue to a neutral residue or a negatively charged residue. In any of these embodiments, the

mutants described may also be made in a *Fok*I domain comprising one or more additional mutations. In preferred embodiments, these additional mutations are in the dimerization domain, e.g. at positions 418, 432, 441, 481, 483, 486, 487, 490, 496, 499, 523, 527, 537, 538 and/or 559. Non-limiting examples of mutations include mutations (e.g., substitutions) of the wild-type residues of any cleavage domain (e.g., *Fok*I or homologue of *Fok*I) at positions 393, 394, 398, 416, 421, 422, 442, 444, 472, 473, 478, 480, 525 or 530 with any amino acid residue (e.g., K393X, K394X, R398X, R416S, D421X, R422X, K444X, S472X, G473X, S472, P478X, G480X, K525X, and A530X, where the first residue depicts wild-type and X refers to any amino acid that is substituted for the wild-type residue). In some embodiments, X is E, D, H, A, K, S, T, D or N. Other exemplary mutations include S418E, S418D, S446D, K448A, I479Q, I479T, Q481A, Q481N, Q481E, A530E and/or A530K wherein the amino acid residues are numbered relative to full length *Fok*I wild-type cleavage domain and homologues thereof. In certain embodiments, combinations may include 416 and 422, a mutation at position 416 and K448A, K448A and I479Q, K448A and Q481A and/or K448A and a mutation at position 525. In one embodiment, the wild-residue at position 416 may be replaced with a Glu (E) residue (R416E), the wild-type residue at position 422 is replaced with a His (H) residue (R422H), and the wild-type residue at position 525 is replaced with an Ala (A) residue. The cleavage domains as described herein can further include additional mutations, including but not limited to at positions 432, 441, 483, 486, 487, 490, 496, 499, 527, 537, 538 and/or 559, for example dimerization domain mutants (e.g., ELD, KKR) and or nickase mutants (mutations to the catalytic domain). The cleavage half-domains with the mutations described herein form heterodimers as known in the art.

[0184] Nucleases may be assembled *in vivo* at the nucleic acid target site using so-called "split-enzyme" technology (see e.g. U.S. Patent Publication No. 2009/0068164). Components of such split enzymes may be expressed either on separate expression constructs, or can be linked in one open reading frame where the individual components are separated, for example, by a self-cleaving 2A peptide or IRES sequence. Components may be individual zinc finger binding domains or domains of a meganuclease nucleic acid binding domain.

[0185] Nucleases (e.g., ZFNs and/or TALENs) can be screened for activity prior to use, for example in a yeast-based chromosomal system as described in as described in U.S. Patent No. 8,563,314.

[0186] In addition to or instead of ZFN or TALEN nickases, the DNA-editing complexes may comprise a CRISPR/Cas nickase. Non-limiting examples of CRISPR/Cas nickases are described in U.S. Patent Nos. 9,840,713; 9,770,489; 9,567,604; 8,932,814; 8,889,356; 8,697,359; 8,771,945; 8,795,965; 8,865,406; 8,871,445; 8,889,356; 8,895,308; 8.906,616; 8,932,814; 8,945,839; 8,945,839; 8,999,641; 10,000,772 and the like.

[0187] The CRISPR (clustered regularly interspaced short palindromic repeats) locus, which encodes RNA components of the system, and the Cas (CRISPR-associated) locus, which encodes proteins (Jansen, *et al.* (2002) *Mol. Microbiol.* 43:1565-1575; Makarova, *et al.* (2002) *Nucleic Acids Res.* 30: 482-496; Makarova, *et al.* (2006) *Biol. Direct* 1:7; Haft, *et al.* (2005) *PLoS Comput. Biol.* 1:e60) make up the gene sequences of the CRISPR/Cas nuclease system. CRISPR loci in microbial hosts contain a combination of CRISPR-associated (Cas) genes as well as non-coding RNA elements capable of programming the specificity of the CRISPR-mediated nucleic acid cleavage.

[0188] The Type II CRISPR is one of the most well characterized systems and carries out targeted DNA double-strand break in four sequential steps. First, two non-coding RNA, the pre-crRNA array and tracrRNA, are transcribed from the CRISPR locus. Second, tracrRNA hybridizes to the repeat regions of the pre-crRNA and mediates the processing of pre-crRNA into mature crRNAs containing individual spacer sequences. Third, the mature crRNA:tracrRNA complex directs Cas9 to the target DNA via Watson-Crick base-pairing between the spacer on the crRNA and the protospacer on the target DNA next to the protospacer adjacent motif (PAM), an additional requirement for target recognition. Finally, Cas9 mediates cleavage of target DNA to create a double-stranded break within the protospacer. Activity of the CRISPR/Cas system comprises of three steps: (i) insertion of alien DNA sequences into the CRISPR array to prevent future attacks, in a process called 'adaptation', (ii) expression of the relevant proteins, as well as expression and processing of the array, followed by (iii) RNA-mediated interference with the alien nucleic acid. Thus, in the bacterial cell, several of the so-called 'Cas' proteins are involved with the natural function of the CRISPR/Cas system and serve roles in functions such as insertion of the alien DNA etc.

[0189] Initially, Cas makes extensive contacts with the ribose–phosphate backbone of the guide RNA, preordering the 10-nt RNA seed sequence required for

initial DNA interrogation. In addition to the preordered seed sequences, the PAM-interacting sites of the Cas protein R1333 and R1335, which are responsible for 5′-NGG-3′ PAM recognition and disordered in the apo structure lacking the guide, are prepositioned prior to making contact with target DNA, indicating that sgRNA loading enables Cas to form a DNA recognition–competent structure. Once Cas binds its guide RNA, the complex is ready to search for complementary target DNA sites. Target search and recognition require both complementary base pairing between the 20-nt spacer sequence and a protospacer in the target DNA, as well as the presence of conserved PAM sequence adjacent to the target site. The PAM sequence is crucial for the discrimination between self and non-self sequences. Single-molecule experiments have demonstrated that Cas initiates the target DNA search process by probing for a proper PAM sequence before interrogating the flanking DNA for potential guide RNA complementarity. Target recognition occurs through three-dimensional collisions, in which Cas rapidly dissociates from DNA that does not contain the appropriate PAM sequence, and dwell time depends on the complementarity between guide RNA and adjacent DNA when a proper PAM is present. Once Cas has found a target site with the appropriate PAM, it triggers local DNA melting at the PAM-adjacent nucleation site, followed by RNA strand invasion to form an RNA–DNA hybrid and a displaced DNA strand (termed R-loop) from PAM-proximal to PAM-distal ends. The PAM duplex is nestled in a positively charged groove between the alpha-helical recognition (REC) lobe the nuclease (NUC) lobe containing the conserved HNH and the split RuvC nuclease domains, with the PAM-containing nontarget strand residing mainly in the C-terminal domain (CTD). The first base in the PAM sequence, denoted as N, remains base paired with its counterpart but does not interact with Cas. The conserved PAM GG dinucleotides are directly read out in the major groove by base-specific hydrogen-bonding interactions with two arginine residues (R1333 and R1335) that are located in a β-hairpin of the CTD. In addition to base-specific contacts with GG dinucleotides, Cas's CTD makes numerous hydrogen-bonding interactions with the deoxyribose-phosphate backbone of the PAM-containing nontarget DNA strand. However, no direct contact has been observed between Cas and target-strand nucleotides complementary to the PAM (Jiang and Doudna (2017) *Annual Review of Biophysics* 46:505-529). In some embodiments, the Cas disclosed in the methods and compositions of the invention is PAM agnostic. In some embodiments, positions R1333 and R1335 as disclosed above comprise

mutations to alter PAM recognition (Anders, *et al.* (2014) *Nature* 513(7519):569-573).

[0190] In some embodiments, the CRISPR-Cpf1 system is used. The CRISPR-Cpf1 system, identified in *Francisella spp.*, is a class 2 CRISPR-Cas system that mediates robust DNA interference in human cells. Although functionally conserved, Cpf1 and Cas9 differ in many aspects including in their guide RNAs and substrate specificity (see Fagerlund, *et al.* (2015) *Genom Bio* 16:251). A major difference between Cas9 and Cpf1 proteins is that Cpf1 does not utilize tracrRNA, and thus requires only a crRNA. The FnCpf1 crRNAs are 42–44 nucleotides long (19-nucleotide repeat and 23–25-nucleotide spacer) and contain a single stem-loop, which tolerates sequence changes that retain secondary structure. In addition, the Cpf1 crRNAs are significantly shorter than the ~100-nucleotide engineered sgRNAs required by Cas9, and the PAM requirements for FnCpf1 are 5′-TTN-3′ and 5′-CTA-3′ on the displaced strand. Although both Cas9 and Cpf1 make double strand breaks in the target DNA, Cas9 uses its RuvC- and HNH-like domains to make blunt-ended cuts within the seed sequence of the guide RNA, whereas Cpf1 uses a RuvC-like domain to produce staggered cuts outside of the seed. Because Cpf1 makes staggered cuts away from the critical seed region, NHEJ will not disrupt the target site, therefore ensuring that Cpf1 can continue to cut the same site until the desired HDR recombination event has taken place. Thus, in the methods and compositions described herein, it is understood that the term "Cas" includes both Cas9 and Cfp1 proteins. Thus, as used herein, a "CRISPR/Cas system" refers both CRISPR/Cas and/or CRISPR/Cfp1 systems, including both nuclease, nickase and/or transcription factor systems.

[0191] In some embodiments, other Cas proteins may be used. Some exemplary Cas proteins include Cas9, Cpf1 (also known as Cas12a), C2c1, C2c2 (also known as Cas13a), C2c3, Cas1, Cas2, Cas4, CasX and CasY; and include engineered and natural variants thereof (Burstein, *et al.* (2017) *Nature* 542:237-241) for example HF1/spCas9 (Kleinstiver, *et al.* (2016) *Nature* 529:490-495; Cebrian-Serrano and Davies (2017) *Mamm Genome* 28(7):247-261); split Cas9 systems (Zetsche, *et al.* (2015) *Nat Biotechnol* 33(2):139-142), trans-spliced Cas9 based on an intein-extein system (Troung, *et al.* (2015) *Nucl Acid Res* 43(13):6450-8); mini-SaCas9 (Ma, *et al.* (2018) *ACS Synth Biol* 7(4):978-985). Thus, in the methods and compositions described herein, it is understood that the term "Cas" includes all Cas variant

proteins, both natural and engineered. Thus, as used herein, a "CRISPR/Cas system" refers to any CRISPR/Cas system, including both nuclease, nickase and/or transcription factor systems.

[0192] In certain embodiments, Cas protein may be a "functional derivative" of a naturally occurring Cas protein. A "functional derivative" of a native sequence polypeptide is a compound having a qualitative biological property in common with a native sequence polypeptide. "Functional derivatives" include, but are not limited to, fragments of a native sequence and derivatives of a native sequence polypeptide and its fragments, provided that they have a biological activity in common with a corresponding native sequence polypeptide. A biological activity contemplated herein is the ability of the functional derivative to hydrolyze a DNA substrate into fragments. The term "derivative" encompasses both amino acid sequence variants of polypeptide, covalent modifications, and fusions thereof such as derivative Cas proteins. Suitable derivatives of a Cas polypeptide or a fragment thereof include but are not limited to mutants, fusions, covalent modifications of Cas protein or a fragment thereof. Cas protein, which includes Cas protein or a fragment thereof, as well as derivatives of Cas protein or a fragment thereof, may be obtainable from a cell or synthesized chemically or by a combination of these two procedures. The cell may be a cell that naturally produces Cas protein, or a cell that naturally produces Cas protein and is genetically engineered to produce the endogenous Cas protein at a higher expression level or to produce a Cas protein from an exogenously introduced nucleic acid, which nucleic acid encodes a Cas that is same or different from the endogenous Cas. In some case, the cell does not naturally produce Cas protein and is genetically engineered to produce a Cas protein. In some embodiments, the Cas protein is a small Cas9 ortholog for delivery via an AAV vector (Ran, *et al.* (2015) *Nature* 510:186).

Delivery

[0193] The DNA-editing complexes (or component molecules thereof) described herein may be delivered to a target cell by any suitable means, including, for example, by injection of the protein and/or mRNA components. Delivery may be to isolated cells (which in turn may be administered to a living subject for *ex vivo* cell therapy) or a living subject via any suitable means. Delivery of gene editing molecules to cells and subjects are known in the art.

[0194] Suitable cells include but not limited to eukaryotic and prokaryotic

cells and/or cell lines. Non-limiting examples of such cells or cell lines generated from such cells include T-cells, COS, CHO (e.g., CHO-S, CHO-K1, CHO-DG44, CHO-DUXB11, CHO-DUKX, CHOK1SV), VERO, MDCK, WI38, V79, B14AF28-G3, BHK, HaK, NS0, SP2/0-Ag14, HeLa, HEK293 (e.g., HEK293-F, HEK293-H, HEK293-T), and perC6 cells as well as insect cells such as *Spodoptera fugiperda* (Sf), or fungal cells such as Saccharomyces, Pichia and Schizosaccharomyces. In certain embodiments, the cell line is a CHO-K1, MDCK or HEK293 cell line. Suitable cells also include stem cells such as, by way of example, embryonic stem cells, induced pluripotent stem cells (iPS cells), hematopoietic stem cells, neuronal stem cells and mesenchymal stem cells.

[0195] Methods of delivering proteins as described herein are described, for example, in U.S. Patent Nos. 6,453,242; 6,503,717; 6,534,261; 6,599,692; 6,607,882; 6,689,558; 6,824,978; 6,933,113; 6,979,539; 7,013,219; and 7,163,824, the disclosures of all of which are incorporated by reference herein in their entireties.

[0196] DNA-editing complexes as described herein may also be delivered using vectors containing sequences encoding one or more of the components (e.g., fusion molecules). Additionally, additional nucleic acids (e.g., donors) also may be delivered via these vectors. Any vector systems may be used including, but not limited to, plasmid vectors, retroviral vectors, lentiviral vectors, adenovirus vectors, poxvirus vectors; herpesvirus vectors and adeno-associated virus vectors, etc. See, also, U.S. Patent Nos. 6,534,261; 6,607,882; 6,824,978; 6,933,113; 6,979,539; 7,013,219; and 7,163,824, incorporated by reference herein in their entireties. Furthermore, it will be apparent that any of these vectors may comprise one or more DNA-binding protein-encoding sequences and/or additional nucleic acids as appropriate. Thus, when one or more DNA-binding proteins as described herein are introduced into the cell, and additional DNAs as appropriate, they may be carried on the same vector or on different vectors. When multiple vectors are used, each vector may comprise a sequence encoding one or multiple DNA-binding proteins and additional nucleic acids as desired. Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids encoding engineered DNA-binding proteins in cells (e.g., mammalian cells) and target tissues and to co-introduce additional nucleotide sequences as desired. Such methods can also be used to administer nucleic acids (e.g., encoding DNA-binding proteins and/or donors) to cells in vitro. In certain embodiments, nucleic acids are administered for in vivo or ex vivo

gene therapy uses. Non-viral vector delivery systems include DNA plasmids, naked nucleic acid, and nucleic acid complexed with a delivery vehicle such as a liposome or poloxamer. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For a review of gene therapy procedures, see Anderson (1992) *Science* 256:808-813; Nabel & Felgner (1993) *TIBTECH* 11:211-217; Mitani & Caskey (1993) *TIBTECH* 11:162-166; Dillon (1993) *TIBTECH* 11:167-175; Miller (1992) *Nature* 357:455-460; Van Brunt (1988) *Biotechnology* 6(10):1149-1154; Vigne (1995) *Restorative Neurology and Neuroscience* 8:35-36; Kremer & Perricaudet (1995) *British Medical Bulletin* 51(1):31-44; Haddada, *et al*, in Current Topics in Microbiology and Immunology Doerfler and Böhm (eds.) (1995); and Yu, *et al.* (1994) *Gene Therapy* 1:13-26.

[0197] Methods of non-viral delivery of nucleic acids include electroporation, lipofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, mRNA, artificial virions, and agent-enhanced uptake of DNA. Sonoporation using, e.g., the Sonitron 2000 system (Rich-Mar) can also be used for delivery of nucleic acids. In a preferred embodiment, one or more nucleic acids are delivered as mRNA. Also preferred is the use of capped mRNAs to increase translational efficiency and/or mRNA stability. Especially preferred are ARCA (anti-reverse cap analog) caps or variants thereof. See U.S. Patent Nos. 7,074,596 and 8,153,773, incorporated by reference herein.

[0198] Additional exemplary nucleic acid delivery systems include those provided by Amaxa Biosystems (Cologne, Germany), Maxcyte, Inc. (Rockville, Maryland), BTX Molecular Delivery Systems (Holliston, MA) and Copernicus Therapeutics Inc, (see for example U.S. Patent No. 6,008,336). Lipofection is described in e.g., U.S. Patent Nos. 5,049,386; 4,946,787; and 4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™, Lipofectin™, and Lipofectamine™ RNAiMAX). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, International Patent Publication Nos. WO 91/17424 and WO 91/16024. Delivery can be to cells (ex vivo administration) or target tissues (in vivo administration).

[0199] The preparation of lipid:nucleic acid complexes, including targeted liposomes such as immunolipid complexes, is well known to one of skill in the art (see, e.g., Crystal (1995) *Science* 270:404-410; Blaese, *et al.* (1995) *Cancer Gene Ther.* 2:291-297; Behr, *et al.* (1994) *Bioconjugate Chem.* 5:382-389; Remy, *et al.*

(1994) *Bioconjugate Chem.* 5:647-654; Gao, *et al.* (1995) *Gene Therapy* 2:710-722; Ahmad, *et al.* (1992) *Cancer Res.* 52:4817-4820; U.S. Patent Nos. 4,186,183; 4,217,344; 4,235,871; 4,261,975; 4,485,054; 4,501,728; 4,774,085; 4,837,028; and 4,946,787).

[0200] Additional methods of delivery include the use of packaging the nucleic acids to be delivered into EnGeneIC delivery vehicles (EDVs). These EDVs are specifically delivered to target tissues using bispecific antibodies where one arm of the antibody has specificity for the target tissue and the other has specificity for the EDV. The antibody brings the EDVs to the target cell surface and then the EDV is brought into the cell by endocytosis. Once in the cell, the contents are released (see MacDiarmid, *et al.* (2009) *Nature Biotechnology* 27(7):643).

[0201] The use of RNA or DNA viral based systems for the delivery of nucleic acids encoding engineered DNA-binding proteins, and/or donors (e.g. CARs or ACTRs) as desired takes advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to patients (in vivo) or they can be used to treat cells in vitro and the modified cells are administered to patients (ex vivo). Conventional viral based systems for the delivery of nucleic acids include, but are not limited to, retroviral, lentivirus, adenoviral, adeno-associated, vaccinia and herpes simplex virus vectors for gene transfer. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

[0202] The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system depends on the target tissue. Retroviral vectors are comprised of cis-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum cis-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian

Immunodeficiency virus (SIV), human immunodeficiency virus (HIV), and combinations thereof (see, e.g., Buchscher, *et al.* (1992) *J. Virol.* 66:2731-2739; Johann, *et al.* (1992) *J. Virol.* 66:1635-1640; Sommerfelt, *et al.* (1990) *Virol.* 176:58-59; Wilson, *et al.* (1989) *J. Virol.* 63:2374-2378; Miller, *et al.* (1991) *J. Virol.* 65:2220-2224; International Patent Publication No. WO 94/26877).

[0203] In applications in which transient expression is preferred, adenoviral based systems can be used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and high levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors are also used to transduce cells with target nucleic acids, e.g., in the in vitro production of nucleic acids and peptides, and for in vivo and ex vivo gene therapy procedures (see, e.g., West, *et al.* (1987) *Virology* 160:38-47; U.S. Patent No. 4,797,368; International Patent Publication No. WO 93/24641; Kotin (1994) *Human Gene Therapy* 5:793-801; Muzyczka (1994) *J. Clin. Invest.* 94:1351. Construction of recombinant AAV vectors are described in a number of publications, including U.S. Patent No. 5,173,414; Tratschin, *et al.* (1985) *Mol. Cell. Biol.* 5:3251-3260; Tratschin, *et al.* (1984) *Mol. Cell. Biol.* 4:2072-2081; Hermonat & Muzyczka (1984) *PNAS USA* 81:6466-6470; and Samulski, *et al.* (1989) *J. Virol.* 63:03822-3828.

[0204] At least six viral vector approaches are currently available for gene transfer in clinical trials, which utilize approaches that involve complementation of defective vectors by genes inserted into helper cell lines to generate the transducing agent.

[0205] pLASN and MFG-S are examples of retroviral vectors that have been used in clinical trials (Dunbar, *et al.* (1995) *Blood* 85:3048-305; Kohn, *et al.* (1995) *Nat. Med.* 1:1017-102; Malech, *et al.* (1997) *PNAS USA* 94(22):12133-12138). PA317/pLASN was the first therapeutic vector used in a gene therapy trial. (Blaese, *et al.* (1995) *Science* 270:475-480). Transduction efficiencies of 50% or greater have been observed for MFG-S packaged vectors. (Ellem, *et al.* (1997) *Immunol Immunother.* 44(1):10-20; Dranoff, *et al.* (1997) *Hum. Gene Ther.* 1:111-2.

[0206] Recombinant adeno-associated virus vectors (rAAV) are a promising alternative gene delivery system based on the defective and nonpathogenic parvovirus adeno-associated type 2 virus. All vectors are derived from a plasmid that retains only the AAV 145 bp inverted terminal repeats flanking the transgene expression

cassette. Efficient gene transfer and stable transgene delivery due to integration into the genomes of the transduced cell are key features for this vector system. (Wagner, *et al.* (1998) *Lancet* 351(9117):1702-3, Kearns, *et al.* (1996) *Gene Ther.* 9:748-55). Other AAV serotypes, including AAV1, AAV3, AAV4, AAV5, AAV6, AAV8, AAV8.2, AAV9 and AAVrh10 and pseudotyped AAV such as AAV2/8, AAV2/5 and AAV2/6 can also be used in accordance with the present invention. Replication-deficient recombinant adenoviral vectors (Ad) can be produced at high titer and readily infect a number of different cell types. Most adenovirus vectors are engineered such that a transgene replaces the Ad E1a, E1b, and/or E3 genes; subsequently the replication defective vector can be propagated in human 293 cells that supply deleted gene function in trans. Ad vectors can transduce multiple types of tissues in vivo, including nondividing, differentiated cells such as those found in liver, kidney and muscle. Conventional Ad vectors have a large carrying capacity. An example of the use of an Ad vector in a clinical trial involved polynucleotide therapy for antitumor immunization with intramuscular injection (Sterman *et al.* (1998) *Hum. Gene Ther.* 7:1083-1089). Additional examples of the use of adenovirus vectors for gene transfer in clinical trials include Rosenecker, *et al.* (1996) *Infection* 24(1):5-10; Sterman, *et al.* (1998) *Hum. Gene Ther.* 9(7):1083-1089; Welsh, *et al.* (1995) *Hum. Gene Ther.* 2:205-218; Alvarez, *et al.* (1997) *Hum. Gene Ther.* 5:597-613; Topf, *et al.* (1998) *Gene Ther.* 5:507-513; Sterman, *et al.* (1998) *Hum. Gene Ther.* 7:1083-1089.

[0207] Packaging cells are used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and ψ2 cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by a producer cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host (if applicable), other viral sequences being replaced by an expression cassette encoding the protein to be expressed. The missing viral functions are supplied in trans by the packaging cell line. For example, AAV vectors used in gene therapy typically only possess inverted terminal repeat (ITR) sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely rep and cap, but lacking ITR sequences. The cell line is also infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and expression of AAV genes

from the helper plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, e.g., heat treatment to which adenovirus is more sensitive than AAV.

[0208] In many gene therapy applications, it is desirable that the gene therapy vector be delivered with a high degree of specificity to a particular tissue type. Accordingly, a viral vector can be modified to have specificity for a given cell type by expressing a ligand as a fusion protein with a viral coat protein on the outer surface of the virus. The ligand is chosen to have affinity for a receptor known to be present on the cell type of interest. For example, Han, *et al.* (1995) *Proc. Natl. Acad. Sci. USA* 92:9747-9751, reported that Moloney murine leukemia virus can be modified to express human heregulin fused to gp70, and the recombinant virus infects certain human breast cancer cells expressing human epidermal growth factor receptor. This principle can be extended to other virus-target cell pairs, in which the target cell expresses a receptor and the virus expresses a fusion protein comprising a ligand for the cell-surface receptor. For example, filamentous phage can be engineered to display antibody fragments (e.g., FAB or Fv) having specific binding affinity for virtually any chosen cellular receptor. Although the above description applies primarily to viral vectors, the same principles can be applied to nonviral vectors. Such vectors can be engineered to contain specific uptake sequences which favor uptake by specific target cells.

[0209] Delivery methods for CRISPR/Cas systems can comprise those methods described above. For example, in animal models, in vitro transcribed Cas encoding mRNA or recombinant Cas protein can be directly injected into one-cell stage embryos using glass needles to genome-edited animals. To express Cas and guide RNAs in cells in vitro, typically plasmids that encode them are transfected into cells via lipofection or electroporation. Also, recombinant Cas protein can be complexed with in vitro transcribed guide RNA where the Cas-guide RNA ribonucleoprotein is taken up by the cells of interest (Kim, *et al.* (2014) *Genome Res* 24(6):1012). For therapeutic purposes, Cas and guide RNAs can be delivered by a combination of viral and non-viral techniques. For example, mRNA encoding Cas may be delivered via nanoparticle delivery while the guide RNAs and any desired transgene or repair template are delivered via AAV (Yin, *et al.* (2016) *Nat Biotechnol* 34(3):328).

[0210] Gene therapy vectors can be delivered in vivo by administration to an

individual patient (subject), typically by systemic administration (e.g., intravenous, intraperitoneal, intramuscular, subdermal, or intracranial infusion) or topical application, as described below. Alternatively, vectors can be delivered to cells ex vivo, such as cells explanted from an individual patient (e.g., lymphocytes, bone marrow aspirates, tissue biopsy) or universal donor hematopoietic stem cells, followed by re-implantation of the cells into a patient, usually after selection for cells which have incorporated the vector.

[0211] Ex vivo cell transfection for diagnostics, research, transplant or for gene therapy (e.g., via re-infusion of the transfected cells into the host organism) is well known to those of skill in the art. In a preferred embodiment, cells are isolated from the subject organism, transfected with a DNA-binding proteins nucleic acid (gene or cDNA), and re-infused back into the subject organism (e.g., patient). Various cell types suitable for ex vivo transfection are well known to those of skill in the art (see, e.g., Freshney, *et al.*, Culture of Animal Cells, A Manual of Basic Technique (3rd ed. 1994)) and the references cited therein for a discussion of how to isolate and culture cells from patients).

[0212] In one embodiment, stem cells are used in ex vivo procedures for cell transfection and gene therapy. The advantage to using stem cells is that they can be differentiated into other cell types in vitro, or can be introduced into a mammal (such as the donor of the cells) where they will engraft in the bone marrow. Methods for differentiating CD34+ cells in vitro into clinically important immune cell types using cytokines such a GM-CSF, IFN-γ and TNF-α are known (see Inaba, *et al.* (1992) *J. Exp. Med.* 176:1693-1702).

[0213] Stem cells are isolated for transduction and differentiation using known methods. For example, stem cells are isolated from bone marrow cells by panning the bone marrow cells with antibodies which bind unwanted cells, such as CD4+ and CD8+ (T cells), CD45+ (panB cells), GR-1 (granulocytes), and Iad (differentiated antigen presenting cells) (see Inaba, *et al.* (1992) *J. Exp. Med.* 176:1693-1702).

[0214] Stem cells that have been modified may also be used in some embodiments. For example, neuronal stem cells that have been made resistant to apoptosis may be used as therapeutic compositions where the stem cells also contain the ZFP TFs of the invention. Resistance to apoptosis may come about, for example, by knocking out BAX and/or BAK using BAX- or BAK-specific ZFNs (see, U.S.

Patent No. 8,597,912) in the stem cells, or those that are disrupted in a caspase, again using caspase-6 specific ZFNs for example.

[0215] Vectors (e.g., retroviruses, adenoviruses, liposomes, etc.) containing therapeutic DNA-binding proteins (or nucleic acids encoding these proteins) can also be administered directly to an organism for transduction of cells in vivo. Alternatively, naked DNA can be administered. Administration is by any of the routes normally used for introducing a molecule into ultimate contact with blood or tissue cells including, but not limited to, injection, infusion, topical application and electroporation. Suitable methods of administering such nucleic acids are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

[0216] Methods for introduction of DNA into hematopoietic stem cells are disclosed, for example, in U.S. Patent No. 5,928,638. Vectors useful for introduction of transgenes into hematopoietic stem cells, e.g., CD34+ cells, include adenovirus Type 35.

[0217] Vectors suitable for introduction of transgenes into immune cells (e.g., T-cells) include non-integrating lentivirus vectors. See, for example, Ory, *et al.* (1996) *Proc. Natl. Acad. Sci. USA* 93:11382-11388; Dull, *et al.* (1998) *J. Virol.* 72:8463-8471; Zuffery, *et al.* (1998) *J. Virol.* 72:9873-9880; Follenzi, *et al.* (2000) *Nature Genetics* 25:217-222.

[0218] Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions available, as described below (see, e.g., Remington's Pharmaceutical Sciences, 17th ed., 1989).

[0219] As noted above, the disclosed methods and compositions can be used in any type of cell including, but not limited to, prokaryotic cells, fungal cells, Archaeal cells, plant cells, insect cells, animal cells, vertebrate cells, mammalian cells and human cells, including T-cells and stem cells of any type. Suitable cell lines for protein expression are known to those of skill in the art and include, but are not limited to COS, CHO (e.g., CHO-S, CHO-K1, CHO-DG44, CHO-DUXB11), VERO, MDCK, WI38, V79, B14AF28-G3, BHK, HaK, NS0, SP2/0-Ag14, HeLa, HEK293 (e.g., HEK293-F, HEK293-H, HEK293-T), perC6, insect cells such as Spodoptera

fugiperda (Sf), and fungal cells such as Saccharomyces, Pichia and Schizosaccharomyces. Progeny, variants and derivatives of these cell lines can also be used.

Applications

[0220] Use of engineered DNA base editor complexes in treatment and prevention of disease provides a significant development in medicine. The methods and compositions described herein serve to increase the specificity of these novel tools to ensure that the desired target sites will be the primary place of editing.

[0221] Exemplary genetic diseases that may be treated and/or prevented by the compositions and methods described herein include, but are not limited to, achondroplasia, achromatopsia, acid maltase deficiency, adenosine deaminase deficiency (OMIM No.102700), adrenoleukodystrophy, aicardi syndrome, alpha-1 antitrypsin deficiency, alpha-thalassemia, androgen insensitivity syndrome, apert syndrome, arrhythmogenic right ventricular, dysplasia, ataxia telangictasia, barth syndrome, beta-thalassemia, blue rubber bleb nevus syndrome, canavan disease, chronic granulomatous diseases (CGD), cri du chat syndrome, cystic fibrosis, dercum's disease, ectodermal dysplasia, fanconi anemia, fibrodysplasia ossificans progressive, fragile X syndrome, galactosemis, Gaucher's disease, generalized gangliosidoses (e.g., GM1), hemochromatosis, the hemoglobin C mutation in the 6th codon of beta-globin (HbC), hemophilia, Huntington's disease, Hurler Syndrome, hypophosphatasia, Klinefleter syndrome, Krabbes Disease, Langer-Giedion Syndrome, leukocyte adhesion deficiency (LAD, OMIM No. 116920), leukodystrophy, long QT syndrome, Marfan syndrome, Moebius syndrome, mucopolysaccharidosis (MPS), nail patella syndrome, nephrogenic diabetes insipdius, neurofibromatosis, Neimann-Pick disease, osteogenesis imperfecta, phenylketonuria (PKU). porphyria, Prader-Willi syndrome, progeria, Proteus syndrome, retinoblastoma, Rett syndrome, Rubinstein-Taybi syndrome, Sanfilippo syndrome, severe combined immunodeficiency (SCID), Shwachman syndrome, sickle cell disease (sickle cell anemia), Smith-Magenis syndrome, Stickler syndrome, Tay-Sachs disease, Thrombocytopenia Absent Radius (TAR) syndrome, Treacher Collins syndrome, trisomy, tuberous sclerosis, Turner's syndrome, urea cycle disorder, von Hippel-Landau disease, Waardenburg syndrome, Williams syndrome, Wilson's disease, Wiskott-Aldrich syndrome, X-linked lymphoproliferative syndrome (XLP,

OMIM No. 308240).

[0222] Additional exemplary diseases that can be treated by targeted DNA base editing include acquired immunodeficiencies, lysosomal storage diseases (e.g., Gaucher's disease, GM1, Fabry disease and Tay-Sachs disease), mucopolysaccahidosis (e.g. Hunter's disease, Hurler's disease), hemoglobinopathies (e.g., sickle cell diseases, HbC, α-thalassemia, β-thalassemia) and hemophilias.

[0223] Such methods also allow for treatment of infections (viral or bacterial) in a host (e.g., by blocking expression of viral or bacterial receptors, thereby preventing infection and/or spread in a host organism) to treat genetic diseases.

[0224] Targeted base editing can also be used to treat viral infections in a host. Additionally, targeted cleavage of genes encoding receptors for viruses can be used to block expression of such receptors, thereby preventing viral infection and/or viral spread in a host organism. Targeted mutagenesis of genes encoding viral receptors (e.g., the CCR5 and CXCR4 receptors for HIV) can be used to render the receptors unable to bind to virus, thereby preventing new infection and blocking the spread of existing infections. See, U.S. Patent Publication No. 2008/015996. Non-limiting examples of viruses or viral receptors that may be targeted include herpes simplex virus (HSV), such as HSV-1 and HSV-2, varicella zoster virus (VZV), Epstein-Barr virus (EBV) and cytomegalovirus (CMV), HHV6 and HHV7. The hepatitis family of viruses includes hepatitis A virus (HAV), hepatitis B virus (HBV), hepatitis C virus (HCV), the delta hepatitis virus (HDV), hepatitis E virus (HEV) and hepatitis G virus (HGV). Other viruses or their receptors may be targeted, including, but not limited to, Picornaviridae (e.g., polioviruses, etc.); Caliciviridae; Togaviridae (e.g., rubella virus, dengue virus, etc.); Flaviviridae; Coronaviridae; Reoviridae; Birnaviridae; Rhabodoviridae (e.g., rabies virus, etc.); Filoviridae; Paramyxoviridae (e.g., mumps virus, measles virus, respiratory syncytial virus, etc.); Orthomyxoviridae (e.g., influenza virus types A, B and C, etc.); Bunyaviridae; Arenaviridae; Retroviradae; lentiviruses (e.g., HTLV-I; HTLV-II; HIV-1 (also known as HTLV-III, LAV, ARV, hTLR, etc.) HIV-II); simian immunodeficiency virus (SIV), human papillomavirus (HPV), influenza virus and the tick-borne encephalitis viruses. See, e.g. Virology, 3rd Edition (W. K. Joklik ed. 1988); Fundamental Virology, 2nd Edition (B. N. Fields and D. M. Knipe, eds. 1991), for a description of these and other viruses. Receptors for HIV, for example, include CCR-5 and CXCR-4. As noted above, the compositions and methods described herein can be used for gene modification, gene correction, and

gene disruption.

[0225] The compositions and methods described herein can also be applied to stem cell based therapies, including but not limited to editing that results in: correction of somatic cell mutations; disruption of dominant negative alleles; disruption of genes required for the entry or productive infection of pathogens into cells; enhanced tissue engineering, for example, by editing gene activity to promote the differentiation or formation of functional tissues; and/or disrupting gene activity to promote the differentiation or formation of functional tissues; blocking or inducing differentiation, for example, by editing genes that block differentiation to promote stem cells to differentiate down a specific lineage pathway Cell types for this procedure include but are not limited to, T-cells, B cells, hematopoietic stem cells, and embryonic stem cells. Additionally, induced pluripotent stem cells (iPSC) may be used which would also be generated from a patient's own somatic cells. Therefore, these stem cells or their derivatives (differentiated cell types or tissues) could be potentially engrafted into any person regardless of their origin or histocompatibility.

[0226] The compositions and methods can also be used for somatic cell therapy, thereby allowing production of stocks of cells that have been modified to enhance their biological properties. Such cells can be infused into a variety of patients, independent of the donor source of the cells and their histocompatibility to the recipient.

[0227] In addition to therapeutic applications, the DNA-editing complexes described herein can be used for crop engineering, cell line engineering and the construction of disease models. The obligate heterodimer cleavage half-domains provide a straightforward means for improving nuclease properties.

[0228] The engineered DNA-editing complexes described can also be used in gene modification protocols requiring simultaneous cleavage at multiple targets at once. Editing at two targets would require cellular expression of two DNA-editing complexes, and is preferably achieved using nickases comprising cleavage domains in each complex that do not interact (dimerize) with cleavage domains in the other complex.

EXAMPLES

Example 1: Preparation of ZFPs

[0229] ZFPs targeted to specific target sites are designed and incorporated into

plasmid vectors essentially as described in Urnov, *et al.* (2005) *Nature* 435(7042):646-651, Perez, *et al.* (2008) *Nature Biotechnology* 26(7): 808-816, and International Patent Publication Nos. WO 2016/183298 and WO 2017/106528. TALEs and sgRNA to specific sites are also developed, as described in U.S. Patent Nos. 8,586,526 and 9,873,894.

[0230] One exemplary target for base editing is the SERPINA locus which encodes Alpha-1 antitrypsin (A1AT). Mutations in the locus that cause an autosomal recessive deficiency in the A1AT protein are associated with both liver and lung disease. The PiZ mutation, one of the most common deficiency alleles in people of Northern European descent, results in only about 10-20% of the A1AT protein being produced. This mutation is caused by a single mutation in exon 5, leading to a glutamine substitution at amino acid position 342 for a lysine where a G at position 1096 in the DNA is an A in the mutated gene sequence (reviewed in Fregonese and Stolk (2008) *Orphanet J Rare Dis* 3:16.

[0231] Another exemplary target is the JAK2 V617F mutation. Editing of the V617F to form another mutation can result in a less activating JAK2. For example, the V617L, V617P and V617S mutations have been shown to be less activating than V617F (Dusa, *et al.* (2008) *J Biol Chem* 283(19):12941-12948).

[0232] Thus, several zinc finger proteins (ZFPs) were made that target the area near the mutations in A1AT (see Figure 2). The design of the ZFPs are shown below in Table 1A (A1AT) and Table 1B (JAK2).

Table 1A: Exemplary ZFP designs for A1AT

SBS #,Target	Design					
	F1	F2	F3	F4	F5	F6
SBS# 78488 atGTTTTTAGAGGC CATacccatgtcta (SEQ ID NO:6)	TQATLGV (SEQ ID NO:11)	DRSHLAR (SEQ ID NO:12)	QSAHRKN (SEQ ID NO:13)	STAALSY (SEQ ID NO:14)	TSGSLTR (SEQ ID NO:15)	N/A
SBS# 78486 atGTTTTTAGAGG CCATacccatgtct a (SEQ ID NO:6)	SNQNLTT (SEQ ID NO:16)	DRSHLAR (SEQ ID NO:12)	QSAHRKN (SEQ ID NO:13)	STAALSY (SEQ ID NO:14)	TSGSLTR (SEQ ID NO:15)	N/A
SBS# 78485 gcCATGTTTTTAGA GGCCATacccatgt (SEQ ID NO:7)	SNQNLTT (SEQ ID NO:16)	DRSHLAR (SEQ ID NO:12)	QSAHRKN (SEQ ID NO:13)	STAALSY (SEQ ID NO:14)	TSGSLSR (SEQ ID NO:17)	HSATLKY (SEQ ID NO:18)
SBS# 78484 gcCATGTTTTTAGA GGCCATacccatgt	SNQNLTT (SEQ ID NO:16)	DRSHLAR (SEQ ID NO:12)	QNAHRKT (SEQ ID NO:19)	STAALSY (SEQ ID NO:14)	TSGSLSR (SEQ ID NO:17)	TSSNRAV (SEQ ID NO:20)

(SEQ ID NO:7)						
SBS# 78483 gcCATGTTTTTAGAGGCCATacccatgt (SEQ ID NO:7)	SNQNLTT (SEQ ID NO:16)	DRSHLAR (SEQ ID NO:12)	QSAHRKN (SEQ ID NO:13)	STAALSY (SEQ ID NO:14)	TSGSLSR (SEQ ID NO:17)	TSSNRAV (SEQ ID NO:20)
SBS# 78482 ggGCCATGttTTTAGAGGCCATacccat (SEQ ID NO:8)	TQATLGV (SEQ ID NO:11)	DRSHLAR (SEQ ID NO:12)	QSAHRKN (SEQ ID NO:13)	STAALSY (SEQ ID NO:14)	RSDALST (SEQ ID NO:21)	DRSTRTK (SEQ ID NO:22)
SBS# 78481 ggGCCATGttTTTAGAGGCCATacccat (SEQ ID NO:8)	SNQNLTT (SEQ ID NO:16)	DRSHLAR (SEQ ID NO:12)	QNAHRKT (SEQ ID NO:19)	STAALSY (SEQ ID NO:14)	RSDALST (SEQ ID NO:21)	DRSTRTK (SEQ ID NO:22)
SBS# 78480 ggGCCATGttTTTAGAGGCCATaccca t(SEQ ID NO:8)	SNQNLTT (SEQ ID NO:16)	DRSHLAR (SEQ ID NO:12)	QSAHRKN (SEQ ID NO:13)	STAALSY (SEQ ID NO:14)	RSDALST (SEQ ID NO:21)	DRSTRTK (SEQ ID NO:22)
SBS# 78477 ggGGCCATGTTTTTAGAggccataccca (SEQ ID NO:9)	QNAHRKT (SEQ ID NO:19)	STAALSY (SEQ ID NO:14)	TSGSLSR (SEQ ID NO:17)	TSSNRAV (SEQ ID NO:20)	DSSHRTR (SEQ ID NO:23)	N/A
SBS# 78476 ggGGCCATGTTTTTAGAggccataccca (SEQ ID NO:9)	QSAHRKN (SEQ ID NO:13)	STAALSY (SEQ ID NO:14)	TSGSLSR (SEQ ID NO:17)	TSSNRAV (SEQ ID NO:20)	DSSHRTR (SEQ ID NO:23)	N/A
SBS# 78475 ctGGGGCCatGTTTTTAGAggccatacc (SEQ ID NO:10)	QSAHRKN (SEQ ID NO:13)	STAALSY (SEQ ID NO:14)	TSGSLTR (SEQ ID NO:15)	DRSDLSR (SEQ ID NO:24)	RSTHLVR (SEQ ID NO:25)	N/A
SBS# 78474 ctGGGGCCatGTTTTTAGAggccatacc (SEQ ID NO:10)	QNAHRKT (SEQ ID NO:19)	STAALSY (SEQ ID NO:14)	TSGSLTR (SEQ ID NO:15)	ERGTLAR (SEQ ID NO:26)	RSDHLSR (SEQ ID NO:27)	N/A

Table 1B: Exemplary ZFP designs for JAK2

ZFN #,Target	Design					
	F1	F2	F3	F4	F5	F6
ZFP2 caAATGCTTGTGAGAAAGCTtgctca (SEQ ID NO:81)	QSSDLSR (SEQ ID NO:82)	LKWNLRT (SEQ ID NO:83)	RSDNLAR (SEQ ID NO:84)	WQSSLIV (SEQ ID NO:85)	QSSDLSR (SEQ ID NO:82)	QSGNRTT (SEQ ID NO:86)
ZFN4 atGCTTGTGAGAAAGCTtgctcatca (SEQ ID NO:87)	QSSDLSR (SEQ ID NO:82)	RSDNLAR (SEQ ID NO:84)	WQSSLIV (SEQ ID NO:85)	QSSDLSR (SEQ ID NO:82)	QSGNRTT (SEQ ID NO:86)	N/A
ZFN6 ctTGTGAGAAAGCTtGCTCATcatac (SEQ ID NO:88)	RSDNLAR (SEQ ID NO:84)	WQSSLIV (SEQ ID NO:85)	QSSDLSR (SEQ ID NO:82)	QSGNRTT (SEQ ID NO:86)	TNQNRIT (SEQ ID NO:89)	RSANLTR (SEQ ID NO:90)
ZFN7 gtGAGAAAGCTtGCTCATcatacttg (SEQ ID NO:91)	AHGARWN (SEQ ID NO:92)	RSANLTR (SEQ ID NO:90)	QSSDLSR (SEQ ID NO:82)	TNQNRIT (SEQ ID NO:89)	RSANLTR (SEQ ID NO:90)	N/A

[0233] Base editors comprising several different combinations are constructed using standard protocols. Different combinations include editors that comprise one or more deaminases, helicases, selected DNA binding domains (ZFP, TALE, sgRNA), dCas, nickase (ZFNs, TALENs, CRISPR/Cas) complexes, UGI, GAM etc. The combinations include complexes comprising the selected domains fused sequentially or wherein the combinations are supplied as separate fusion proteins. Any linker is used between the domains.

[0234] The combinations are assembled into expression constructs for transfection into cells or for use in the production of mRNAs in vitro. These mRNAs can then be introduced into the cells by methods known in the art (*e.g.*, electroporation). Cells without the targeted mutation are used as controls.

Example 2: Adenine Base Editors

[0235] Adenine base editors were constructed using Cas9 variants with relaxed PAM requirements (SpCas9VRVRFRR_D10A; Nishimasu, *et al.* (2018) *Science* 361:1259-1262 or xCas9, see Hu, *et al.* (2018) *Nature* 556:57-63) linked to a ZFP DNA binding domain that targeted the mutated A1AT PiZ mutation on the C-terminal side of the molecule. A series of ZFP DNA binding domains (see Example 1) that were designed to bind to the adjacent DNA region were incorporated into the base editor (see Figures 2 and 7). The ZFP was attached to the Cas9 using a linker comprising 3 HA peptides and two nuclease localization sequences (NLS). The sequence of the linker was: GTGGPKKKRKVYPYDVPDYAGYPYDVPDYAGSYPYDVPDYAGSAAPAAKK KKLDFESE (SEQ ID NO:3) (see Bolukbasi, *et al.* (2015) *Nat Methods* 12(12):1150-1156). The Cas9 was then linked to two *E. coli* TadA adenine deaminases ("ecTadA"; Kim, *et al.* (2006) *Biochemistry* 45:6407-6416) on the N terminal side. In the construct, one of the TadA proteins was a wild type protein while the other was an evolved version (Gaudelli, *et al.* (2017) *Nature* 551:464). A serine-glycine rich linker was used twice between the Cas9 and TadA subunits and comprised the sequence: SGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID NO:2). A schematic of the adenine base editor (ABE) is shown in Figure 3A. In some instances, other types of adenine deaminases are used. For example, in some constructs, the ABE7.10 deaminase or the ABEmax adenine deaminase (Koblan, *et al.* (2018) *Nat Biotechnol.*

36(9):843-846) is used.

[0236] To target the JAK2 V617F mutation, Cas9NG adenine base editors targeting the JAK2 gene near the mutation site are made using the TAA or AAT or AAA PAM sites located upstream of the mutation site (see, Figures 7A through 7C). ZFP DNA binding domains are fused to the base editor as described above and used to edit HSC/PC. ZFP sequences are shown in Table 1B. Any linkers can be used between the fingers and/or between the ZFP and the base editor.

[0237] The results show successful editing at the target locus. Figure 7A shows base editing to cause a change in the amino acid sequence to a serine (S) or a proline (P). Both of these variants have been shown to be less activating than the phenylalanine mutant (F) (see Dusa, *et al.* (2008) *J Biol Chem* 283(19):12941-8).

[0238] Cells (*e.g.*, K562 cells) are transfected with expression vectors (for example plasmids or viral vectors) comprising the base editors, or the cells are electroporated using the mRNAs encoding the base editors as described above. Transfected cells are harvested and the genomic DNA is isolated. On-target and off-target genomic regions of interest are amplified by PCR amplification according to the standard methods in the art. Sequences are evaluated for base editing and for the presence of indels.

[0239] The adenine base editor (ABE) was tested on the A1AT locus as described above on K562 cells where 800ng of plasmid DNA encoding the base editor was used per 200,000 cells. Experiments were also performed in K562 cells without the Z and that A nucleotides in close proximity to the disease-causing mutations were used as a proxy to measure activity. 72 hours after transfection using an Amaxa device (according to manufacturer's protocols), cells were harvested and subject to Miseq analysis (Illumina) to analyze any editing that may have occurred. Figure 4 illustrates the A bases that could be targeted within the editing window of the complex, where the presence of a G at these positions would indicate that the A based editing had occurred.

[0240] Three different clones of each construct comprising the variant ZFP domains were tested. The results are shown below in Table 2 (note that the terms "Cas9VRVRFRR", "Cas9VR" and "Cas9NG" are used interchangeably):

Table 2: Adenine base editing at A1AT locus

	A1 %G	A5 %G	A8 %G	A10 %G	A11 %G	A12 %G	A16 %G	A20 %G
sgRNA_only_r1	0.03	0.12	0.05	0.11	0.06	0.22	0.11	0.03
sgRNA_only_r2	0.08	0.08	0.07	0.09	0.07	0.15	0.14	0.05
sgRNA_only_r3	0.07	0.07	0.06	0.09	0.08	0.06	0.23	0.04
sgRNA_only_r4	0.05	0.04	0.08	0.06	0.06	0.24	0.17	0.02
ABE_Cas9VR_D10A_r1	0.07	0.51	0.10	0.04	0.07	0.22	0.17	0.05
ABE_Cas9VR_D10A_r2	0.04	0.45	0.12	0.06	0.03	0.20	0.13	0.04
ABE_Cas9VR_D10A_r3	0.08	0.53	0.07	0.08	0.03	0.20	0.15	0.04
ABE_Cas9VR_D10A_r4	0.04	0.56	0.13	0.04	0.02	0.17	0.14	0.03
ABE_Cas9VR_D10A_SGMO_SBS_78474 (1)	0.06	2.61	0.14	0.06	0.06	0.21	0.12	0.09
ABE_Cas9VR_D10A_SGMO_SBS_78474 (2)	0.04	2.83	0.19	0.19	0.11	0.28	0.15	0.05
ABE_Cas9VR_D10A_SGMO_SBS_78474 (3)	NA	NA	NA	NA	NA	NA	NA	NA
ABE_Cas9VR_D10A_SGMO_SBS_78475 (1)	0.05	0.29	0.06	0.07	0.02	0.27	0.17	0.06
ABE_Cas9VR_D10A_SGMO_SBS_78475 (2)	0.10	0.08	0.13	0.03	0.10	0.18	0.21	0.07
ABE_Cas9VR_D10A_SGMO_SBS_78475 (3)	0.07	0.41	0.15	0.06	0.05	0.19	0.13	0.10
ABE_Cas9VR_D10A_SGMO_SBS_78476 (1)	0.07	1.10	0.07	0.12	0.02	0.15	0.12	0.05
ABE_Cas9VR_D10A_SGMO_SBS_78476 (2)	0.05	1.22	0.15	0.08	0.05	0.15	0.19	0.11
ABE_Cas9VR_D10A_SGMO_SBS_78476 (3)	0.06	0.83	0.12	0.07	0.06	0.14	0.18	0.05
ABE_Cas9VR_D10A_SGMO_SBS_78477 (1)	0.08	1.50	0.08	0.04	0.03	0.19	0.10	0.03
ABE_Cas9VR_D10A_SGMO_SBS_78477 (2)	0.07	1.40	0.20	0.07	0.08	0.16	0.19	0.11
ABE_Cas9VR_D10A_SGMO_SBS_78477 (3)	0.06	1.03	0.11	0.10	0.06	0.24	0.17	0.01
ABE_Cas9VR_D10A_SGMO_SBS_78480 (1)	0.03	1.64	0.13	0.06	0.04	0.13	0.14	0.07
ABE_Cas9VR_D10A_SGMO_SBS_78480 (2)	0.09	1.61	0.12	0.04	0.08	0.26	0.12	0.03
ABE_Cas9VR_D10A_SGMO_SBS_78480 (3)	0.08	1.42	0.10	0.08	0.09	0.25	0.12	0.06
ABE_Cas9VR_D10A_SGMO_SBS_78481 (1)	NA	NA	NA	NA	NA	NA	NA	NA
ABE_Cas9VR_D10A_SGMO_SBS_78481 (2)	0.06	3.91	0.07	0.08	0.08	0.27	0.15	0.03
ABE_Cas9VR_D10A_SGMO_SBS_78481 (3)	0.06	3.05	0.10	0.12	0.07	0.27	0.14	0.06
ABE_Cas9VR_D10A_SGMO_SBS_78482 (1)	0.07	0.93	0.14	0.08	0.03	0.26	0.21	0.06
ABE_Cas9VR_D10A_SGMO_SBS_78482 (2)	0.05	1.14	0.12	0.12	0.06	0.20	0.21	0.08
ABE_Cas9VR_D10A_SGMO_SBS_78482 (3)	NA	NA	NA	NA	NA	NA	NA	NA

ABE_Cas9VR_D10A_SGMO_SBS_78483 (1)	NA	NA	NA	NA	NA	NA	NA	NA
ABE_Cas9VR_D10A_SGMO_SBS_78483 (2)	0.08	1.25	0.09	0.08	0.03	0.20	0.19	0.05
ABE_Cas9VR_D10A_SGMO_SBS_78483 (3)	0.09	0.86	0.14	0.08	0.09	0.18	0.17	0.04
ABE_Cas9VR_D10A_SGMO_SBS_78484 (1)	0.07	2.48	0.05	0.10	0.07	0.24	0.17	0.06
ABE_Cas9VR_D10A_SGMO_SBS_78484 (2)	0.08	2.47	0.15	0.06	0.04	0.22	0.14	0.04
ABE_Cas9VR_D10A_SGMO_SBS_78484 (3)	0.00	2.06	0.13	0.07	0.09	0.27	0.11	0.03
ABE_Cas9VR_D10A_SGMO_SBS_78485 (1)	0.08	1.08	0.06	0.10	0.04	0.21	0.15	0.03
ABE_Cas9VR_D10A_SGMO_SBS_78485 (2)	0.07	1.05	0.11	0.10	0.02	0.25	0.12	0.06
ABE_Cas9VR_D10A_SGMO_SBS_78485 (3)	0.05	0.61	0.11	0.05	0.03	0.16	0.18	0.04
ABE_Cas9VR_D10A_SGMO_SBS_78486 (1)	NA	NA	NA	NA	NA	NA	NA	NA
ABE_Cas9VR_D10A_SGMO_SBS_78486 (2)	0.09	2.33	0.20	0.10	0.07	0.23	0.18	0.05
ABE_Cas9VR_D10A_SGMO_SBS_78486 (3)	0.11	2.03	0.14	0.12	0.05	0.13	0.19	0.06
ABE_Cas9VR_D10A_SGMO_SBS_78488 (1)	NA	NA	NA	NA	NA	NA	NA	NA
ABE_Cas9VR_D10A_SGMO_SBS_78488 (2)	0.04	1.67	0.12	0.08	0.06	0.22	0.13	0.07
ABE_Cas9VR_D10A_SGMO_SBS_78488 (3)	0.10	1.32	0.08	0.07	0.09	0.23	0.18	0.02
(1), (2) and (3) represent individual clones. NA = Clone did not pass Sanger sequencing QC and was not tested.								

[0241] Rows labeled "sgRNA_only" are those comprising guide RNA only, "ABE_Cas9VR_D10A" are the complexes lacking the ZFP DNA binding domain. As can be seen from the data, targeted editing of the adenine in position A5 increased in the presence of some of the ZFP DNA binding domains as compared to the editing complex lacking the ZFP domain. The native ABE-Cas9 fusion construct without ZFP resulted in ~0.5% base editing while ABE-Cas9-ZFP fusion constructs showed at least 7-fold higher base editing efficiencies in this dataset.

[0242] In these experiments, studies were done using the xCas or the Cas9NG proteins linked to the ABEmax or the ABE7.10 adenine deaminases. The guide RNAs used were either the TGT PAM or the AGT PAM. Results showed base editing efficiencies of 5 to 10-fold or more as compared to base editors lacking the ZFP DNA binding domain (ZFP anchor).

[0243] Further experiments were carried out using alternate versions of the

adenine deaminases including use of ABE 7.9 and ABE 7.8 and showed comparable results.

[0244] For base editing of the JAK2 V617F mutant, experimental conditions were the same as described above, including experiments performed in K562 cells without the V617F mutation but A nucleotides in close proximity to the disease-causing mutations were used as a proxy to measure activity. Targeted PAM sequences are shown in Figure 7B. Studies were done using various combinations of the xCas or the Cas9NG proteins linked to ABEmax (*see, e.g.,* Figure 3B).

[0245] Results showed that the presence of the ZFP anchor improved editing and relaxed the PAM requirements, including showing activity at AAT and TAA PAM sequences. Notably, base editors with TAA PAM sequences are inactive at this site without the ZFP domain. *See, e.g.,* exemplary results as shown in Figure 7C.

[0246] Additional experiments are performed for additional disease-related point mutations and higher base editing specificity and/or activity is achieved in the presence of a ZFP anchor domain. Furthermore, depending on the targeted base to be edited, any PAM sequence can be used, including but not limited to NAN (*e.g.,* TAA), AAT, NGG (TGG), NGT (*e.g.,* TGT or AGT).

Example 3: Cytidine base editors

[0247] A cytidine base editor to convert C nucleotides to U is constructed using the apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC) protein (Yang, *et al.* (2017) *J Genet Genomics* 44(9):423-437). In particular, cytidine base editors are made using a cytidine deaminase such as Apolipoprotein B mRNA editing enzyme, catalytic polypeptide 1 or APOBEC-1. The Activation-induced cytidine deaminase, also known as AICDA and AID, encoded by the AICDA gene, may also be included in cytidine deaminase base editors as described herein.

[0248] It is thought that the cellular repair response to U:G heteroduplex DNA invokes the activity of a uracil DNA glycosylase (UDG) that catalyzes the removal of U from DNA and initiates base-excision repair with reversion of the U:G pair to a C:G pair, decreasing the efficiency of base editing. Thus, cytidine base editors can have a uracil glycosylase inhibitor (UGI) fused to the editor to block endogenous UDG activity (Komor, *et al.* (2016) *Nature* 533(7603):420-424). Cytidine base editors are constructed using a ZFP DNA binding domain linked to an APOBEC1 or

B-cell-specific activation-induced cytidine deaminase (AID) enzyme (Kescu and Adli (2016) *Nat Methods* 13(12):983). The cytidine base editors are attached to a DNA binding domain such as ZFP.

[0249] In some embodiments, the editors further comprise a UGI dimer (see, Figures 5A through 5F). The UGI dimer is attached to the ZFP DNA binding domain using a linker such as a L0 (LRGS, SEQ ID NO:76) or N7a (SGTPHEVGVYTL, SEQ ID NO:28, see U.S. Patent Publication No. 2017/0218349). The AID or APOBEC1 is attached to the ZFP via a linker such as L0 or a sequence such as SGGGLGST (SEQ ID NO:29, Yang, *et al.* (2016) *Nat Commun* DOI:10.1038/ncomms13330).

[0250] Cytidine base editors are also constructed without the UGI domains. Some editors are constructed to be used as a pair where one partner comprises a cytidine editor linked to a ZFP linked to a catalytically inactive *Fok*I nickase domain. The second partner comprises the active Fok domain such that the two Fok half domains can pair and act to create a nick. The cytidine editor domain can be on either half as can a UGI dimer assembly.

Example 4: Additional base editors

[0251] Additional ABE and/or CBE are constructed and tested.

[0252] In particular, experiments were performed using an adenine base editor comprising (1) a Cas9 nickase, optionally operably linked to a ZFP anchor; and (2) a ZFP operably linked to an ABE domain (*e.g.*, evolved ABE domain). *See, e.g.*, Figure 1B, bottom middle panel.

[0253] Results showed that these base editors were effective in targeted editing of the disease-related mutation.

[0254] In addition, experiments are performed with an ABE base editor comprising: (1) a dCas9 protein operably linked to a single guide RNA, optionally operably linked to a ZFP anchor; (2) a ZFP operably linked to an ABE domain (*e.g.*, evolved ABE domain); and (3) a ZFN nickase. *See, e.g.*, Figure 1C.

[0255] Results show that base editors including a ZFN nickase increased base editing efficiency as compared to dCas9 base editors.

Example 5: Cas9-Free Base Editors

[0256] Cas9-free base editors are also constructed and evaluated. Constructs

include base editors with two TadA domain are used as described above, optionally where one is wild type and one is evolved, and these may be linked to a ZFP DNA binding domain. This assembly is used alone, or may then can be linked to a catalytically inactive *Fok*I nickase domain. When used in combination with another vector comprising an active *Fok*I domain, the adenine base editor has nicking activity to prevent correction of the base edit. Other non-Cas base editors made are shown in Figure 1D.

[0257] A non-Cas base editor comprising a DNA destabilizing or unwinding factor is also constructed and tested above. The DNA destabilizing factor is fused to the N- or C-terminus of a ZFP and/or ZFN nickase (see Figure 1D) or introduced independently from the ZFP and/or nickase.

[0258] In particular, base editors are shown in Figure 1D are generated. These base editors include a ZFP-deaminase fusion protein and a ZFN nickase. In addition, these editors include a DNA-destabilizing factor, optionally linked by its 3' end or 5' end to a ZFP of the ZFN nickase.

[0259] In particular, base editors are constructed that include one or more protein DNA-destabilizing factors as shown in Table A (*e.g.*, helicases; factors involved in D-loop formation during DSB repair (*e.g.* Rad51, Rad52, RPA1, RPA2, RPA3, etc.); and/or helix-destabilizing proteins (*e.g.* ICP8, Puralpha or calf-thymus DNA helix-destabilizing protein), with or without one or more CRISPR proteins (*e.g.*, non-Cas9 proteins)).

[0260] Alternatively, or in addition to, the DNA-destabilizing proteins, base editors are constructed that include one or more peptide nucleic acids (PNAs); locked nucleic acids (LNAs) and/or bridged nucleic acids (BNA). In particular, base editors comprising one or more nucleotides are constructed and tested. Base editors comprising PNAs and/or LNAs are constructed as described herein (*see, also,* Figure 1D and Figure 8).

[0261] Results show that Cas9-free base editors edit their target sites.

[0262] All patents, patent applications and publications mentioned herein are hereby incorporated by reference in their entirety.

[0263] Although disclosure has been provided in some detail by way of illustration and example for the purposes of clarity of understanding, it will be

apparent to those skilled in the art that various changes and modifications can be practiced without departing from the spirit or scope of the disclosure. Accordingly, the foregoing descriptions and examples should not be construed as limiting.

CLAIMS

What is claimed is:

1. One or more expression vectors for expressing a system for editing an adenine or cytosine base in DNA, the system comprising:

(i) an adenosine or cytidine deaminase, wherein the deaminase is a dimer comprising

a first fusion protein comprising a first inactive domain of the deaminase and a first ZFP DNA-binding domain, and

a second fusion protein comprising a second inactive domain of the deaminase and a second ZFP DNA-binding domain, wherein the first and second inactive domains dimerize to form an active deaminase; and

(ii) a nickase fused to one or more ZFP DNA-binding domains.

2. One or more expression vectors for expressing a system for editing an adenine or cytosine base in DNA, the system comprising:

(i) an adenosine or cytidine deaminase fused to one or more zinc finger protein (ZFP) DNA-binding domains, and

(ii) a zinc finger nuclease (ZFN) nickase, wherein the ZFN nickase is a dimer comprising

a first polypeptide comprising a catalytically active nickase domain fused to a ZFP DNA-binding domain and

a second polypeptide comprising a catalytically inactive nickase domain fused to another ZFP DNA-binding domain, wherein the first and second polypeptides dimerize to form an active ZFN nickase.

3. The one or more expression vectors of claim 1, wherein

the second fusion protein comprises a catalytically active or inactive ZFN nickase domain, and

the system further comprises a third fusion protein comprising a catalytically inactive or active ZFN nickase domain and a third ZFP DNA-binding domain,

wherein the second and third fusion proteins form an active ZFN nickase through dimerization of the catalytically active and inactive ZFN nickase domains.

4. The one or more expression vectors of any one of claims 1-3, wherein the system further comprises a uracil DNA glycosylase inhibitor (UGI), optionally wherein the UGI is at the N- or C-terminal end of the fusion protein(s).

5. The one or more expression vectors of any one of claims 1-4, wherein the expression vector(s) are viral vectors, plasmid vectors(s), or mRNA vector(s).

6. The one or more expression vectors of claim 5, wherein the expression vector(s) is an adeno-associated viral (AAV) vector.

7. A kit for use in editing a base in DNA, the kit comprising the one or more expression vectors of any one of claims 1-6, optionally wherein the one or more expression vectors are one or more plasmid, viral, or mRNA vectors.

8. A cell comprising the one or more expression vectors of any one of claims 1-6, or a descendant of the cell wherein the cell comprises an edited base.

9. The one or more expression vectors of any one of claims 1-6 for use in a method of editing a base in a target DNA in a cell, the method comprising providing to the cell said one or more expression vectors.

10. The one or more expression vectors for use of claim 9, wherein

(A) the editing comprises:

(i) editing a cytosine base to a uracil base, optionally wherein the uracil base is replaced with a thymine base during DNA replication, or

(ii) editing an adenine base to a hypoxanthine base, optionally wherein the hypoxanthine base is replaced with a guanine base during replication; or

(B) the editing in the cell results in:

(i) changing a C:G base pair to a T:A base pair,

(ii) changing an A:T base pair to a G:C base pair,

(iii) introduction of a stop codon, and/or

(iv) editing or creating a splicing sequence.

11. The one or more expression vectors for use of claim 9 or 10, wherein the editing corrects a disease mutation, optionally wherein an exon is edited.

12. The one or more expression vectors for use of any one of claims 9-11, wherein the base is in a chromosome or an extrachromosomal episome in the cell, optionally wherein the cell is a cell in or obtained from a subject in need of the editing.

13. Use of the cell of claim 8 or a descendant thereof in the manufacture of a medicament for treating a disease in a subject in need thereof, wherein said cell or descendant thereof is to be administered to the subject.

14. The cell of claim 8 or a descendant thereof for use in therapy.