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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 March 2022 (31.03.2022)

WIPO | PCT

(10) International Publication Number
WO 2022/067122 A1

(51) International Patent Classification:
C12N 9/78 (2006.01) *C12N 15/10* (2006.01)

(21) International Application Number:
PCT/US2021/052088

(22) International Filing Date:
24 September 2021 (24.09.2021)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

63/083,662	25 September 2020 (25.09.2020)	US
63/164,893	23 March 2021 (23.03.2021)	US
63/230,580	06 August 2021 (06.08.2021)	US

(71) Applicant: **SANGAMO THERAPEUTICS, INC.** [US/US]; 7000 Marina Boulevard, Brisbane, California 94005 (US).

(72) Inventors: **FAUSER, Friedrich A.**; c/o Sangamo Therapeutics, Inc., 7000 Marina Boulevard, Brisbane, California 94005 (US). **MILLER, Jeffrey C.**; c/o Sangamo Therapeutics, Inc., 7000 Marina Boulevard, Brisbane, California 94005 (US). **ARANGUNDY, Sebastian**; c/o Sangamo Therapeutics, Inc., 7000 Marina Boulevard, Brisbane, California 94005 (US).

(74) Agent: **LI, Z. Ying** et al.; Steptoe & Johnson LLP, 1330 Connecticut Avenue, N.W., Washington, District of Columbia 20036 (US).

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

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

Declarations under Rule 4.17:
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*
— *of inventorship (Rule 4.17(iv))*

Published:
— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) **Title:** ZINC FINGER FUSION PROTEINS FOR NUCLEOBASE EDITING

(57) **Abstract:** Provided herein are base editor systems comprising fusion proteins that comprise zinc finger protein and cytidine deaminase domains, as well as methods of using the base editor systems. The systems can be used to specifically alter a single base pair in a target DNA sequence.

WO 2022/067122 A1

ZINC FINGER FUSION PROTEINS FOR NUCLEOBASE EDITING

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority from United States Provisional Patent Application 63/083,662, filed September 25, 2020; United States Provisional Patent Application 63/164,893, filed March 23, 2021; and United States Provisional Patent Application 63/230,580, filed August 6, 2021. The disclosures of those priority applications are incorporated by reference herein in their entirety.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. The electronic copy of the Sequence Listing, created on September 22, 2021, is named 025297_WO034_SL.txt and is 529,443 bytes in size.

BACKGROUND OF THE INVENTION

[001] Precision DNA editing of single bases has various applications in treating and understanding disorders such as genetic diseases. For example, knock-out of one or more genes can be achieved by converting regular codons into stop codons, or by mutating splice acceptor sites to introduce exon skipping and/or frameshift mutations. Further, DNA point mutations are associated with a wide range of disorders. Single base editing can be used to correct deleterious mutations or to introduce beneficial genetic modifications.

[002] Cytidine deaminases convert the nucleobase cytosine to thymine (or the nucleoside deoxycytidine to thymidine). These enzymes function in the pyrimidine salvage pathway, predominantly operating on single-stranded DNA to convert cytosine into uracil, which is subsequently replaced by a thymine base during DNA replication or repair. A cytidine deaminase identified in the bacterium *Burkholderia cenocepacia*, DddA, can catalyze the deamination of cytosine to uracil within double-stranded DNA. DddA thus bypasses the requirement for unwinding of the dsDNA to ssDNA (Mok et al., *Nature* (2020) 583:631-7). While the Mok study reports C to T base editing at the human *CCR5* locus with a DddA-derived cytosine base editor fused to transcription activator-like effector (TALE) proteins, it

is unclear how broadly this approach is applicable. Further, new deaminases that operate on double-stranded DNA may have improved or altered base editing activity compared to DddA.

[003] Thus, there continues to be a need to develop precise base editing systems for the prevention and treatment of numerous diseases.

SUMMARY OF THE INVENTION

[004] The present disclosure provides zinc finger protein (ZFP) based nucleobase editing systems and uses thereof. In one aspect, the present disclosure provides a system for changing a cytosine to a thymine in the genome of a cell (e.g., a eukaryotic cell or a prokaryotic cell, wherein the eukaryotic cell may be a mammalian cell such as a human cell, or a plant cell), comprising a first fusion protein and a second fusion protein, or first and second expression constructs for expressing the first and second fusion proteins, respectively, wherein a) the first fusion protein comprises: i) a first zinc finger protein (ZFP) domain that binds to a first sequence in a target genomic region in the cell, and ii) a first portion of a cytidine deaminase polypeptide (e.g., wherein the cytidine deaminase is a toxin-derived deaminase (TDD) comprising an amino acid sequence at least 90% identical to SEQ ID NO: 49, 81, 92, 95, 98, 101, 104, 107, 134, 143, 152, 157, 162, 167, 172, 177, 184, 189, 194, 199, 204, 209, 214, or 219); b) the second fusion protein comprises: i) a second ZFP domain that binds to a second sequence in the target genomic region, and ii) a second portion of the cytidine deaminase polypeptide; and c) binding of the first fusion protein and the second fusion protein to the target genomic region results in dimerization of the first and second portions, wherein the dimerized portions form an active cytidine deaminase capable of changing a cytosine to a uracil in the target genomic region. In some embodiments, the first and second portions lack cytidine deaminase activity on their own. In some embodiments, the first and second portions form an active cytidine deaminase that comprises an amino acid sequence at least 90% identical to SEQ ID NO: 49, 81, 92, 95, 98, 101, 104, 107, 134, 143, 152, 157, 162, 167, 172, 177, 184, 189, 194, 199, 204, 209, 214, or 219. In some embodiments, the first and second portions form an active cytidine deaminase that comprises the amino acid sequence of SEQ ID NO: 49, 81, 92, 95, 98, 101, 104, 107, 134, 143, 152, 157, 162, 167, 172, 177, 184, 189, 194, 199, 204, 209, 214, or 219. In some embodiments, the target genomic region may be specific to a particular allele of a gene in the cell. In some embodiments, the targeted cytosine may be between the proximal ends of the first sequence and the second sequence in the target genomic region, optionally wherein the proximal ends are no more than 100 bps apart.

[005] Also provided are multiplex versions of the present base editor systems comprising more than one pair of the first and second fusion proteins, wherein each pair of the fusion proteins binds to a different target genomic region, optionally wherein the first and second cytidine deaminase portions of one pair of fusion proteins are different from the first and second portions of another pair of fusion proteins.

[006] In some embodiments, the base editor system further comprises a nickase that creates a single-stranded DNA break on the unedited or edited strand, wherein the DNA break is no more than about 500 bps, optionally no more than 200 bps, optionally about 10-50 bps, from the cytosine to be edited. The nickase may be, e.g., a ZFP-based nickase, a TALE-based nickase, or a CRISPR-based nickase. In some embodiments, the nickase is a ZFP-based nickase formed by dimerization of a first nickase domain and a second nickase domain fused respectively to two ZFP domains that bind to the target genomic region, wherein the first and second nickase domains are inactive, or lack significant or specific nickase activity, on their own. In certain embodiments, one of the nickase domains is fused to the first or second ZFP-cytidine deaminase fusion protein, and the other nickase domain is fused to a third ZFP domain that binds to a third sequence in the target genomic region. Alternatively, the two nickase domains may be fused respectively to a third ZFP domain that binds a third sequence in the target genomic region and a fourth ZFP domain that binds a fourth sequence in the target genomic region. In particular embodiments, the first and second nickase domains are derived from FokI.

[007] In some embodiments, the base editor system further comprises an inhibitory component of the cytidine deaminase, e.g., a toxin-derived deaminase inhibitor (TDDI) where the cytidine deaminase is a TDD. For example, the inhibitor may be a DddI component where the cytidine deaminase is DddA. In certain embodiments, this system comprises a third fusion protein or a third expression construct for expressing the third fusion protein in the cell, wherein the third fusion protein comprises i) a ZFP domain that binds to a third sequence in the target genomic region, and ii) an inhibitory domain for the cytidine deaminase (e.g., a TDDI where the cytidine deaminase is a TDD, such as DddI where the cytidine deaminase is DddA), and binding of the third fusion protein to the target genomic region results in the interaction of the inhibitory domain with, and thereby inhibition of the cytidine deaminase activity of, the dimerized cytidine deaminase portions.

[008] In some embodiments of the inhibitory domain-containing base editor system, the system comprises a third fusion protein or a third expression construct for expressing the third fusion protein in the cell, and a fourth fusion protein or a fourth expression construct for

expressing the fourth fusion protein in the cell, wherein the third fusion protein comprises i) a ZFP domain that binds to a third sequence in the target genomic region, and ii) a first dimerization domain; and the fourth fusion protein comprises i) an inhibitory domain for the cytidine deaminase (e.g., a TDDI where the cytidine deaminase is a TDD, such as DddI where the cytidine deaminase is DddA), and ii) a second dimerization domain capable of partnering with the first dimerization domain in the presence of a dimerization-inducing agent; and binding of the third fusion protein to the target genomic region and dimerization of the third and fourth fusion proteins result in the binding of the inhibitory domain to, and thereby inhibition of the cytidine deaminase activity of, the dimerized cytidine deaminase portions.

[009] In some embodiments of the inhibitory domain-containing base editor system, the system comprises a third fusion protein or a third expression construct for expressing the third fusion protein in the cell, and a fourth fusion protein or a fourth expression construct for expressing the fourth fusion protein in the cell, wherein the third fusion protein comprises i) a ZFP domain that binds to a third sequence in the target genomic region, and ii) a first dimerization domain; and the fourth fusion protein comprises i) an inhibitory domain for the cytidine deaminase (e.g., a TDDI where the cytidine deaminase is a TDD, such as DddI where the cytidine deaminase is DddA), and ii) a second dimerization domain capable of partnering with the first dimerization domain in the absence of a dimerization-inhibiting agent; and binding of the third fusion protein to the target genomic region, and dimerization of the third and fourth fusion proteins, result in the binding of the inhibitory domain to, and thereby inhibition of the cytidine deaminase activity of, the dimerized cytidine deaminase portions.

[0010] In particular embodiments, the base editor systems described herein comprise both a nickase component and an inhibitory domain component described herein.

[0011] Any of the ZFP domains used in the fusion proteins described herein may independently have 2, 3, 4, 5, 6, 7, or 8 zinc fingers.

[0012] In some embodiments, the protein components of the present base editor systems are provided to the cells by means of expression cassettes or constructs. Such cassettes or constructs may be provided to the cells on the same or separate expression vectors such as viral vectors. The viral vectors may be, e.g., adeno-associated viral (AAV) vectors, adenoviral vectors, or lentiviral vectors.

[0013] In some embodiments of the base editor systems described herein, the cytidine deaminase is a TDD. In certain embodiments, the TDD comprises the amino acid sequence

of SEQ ID NO: 72 (DddA), or the toxic domain of a TDD comprising said sequence (e.g., the toxic domain of SEQ ID NO: 49 or 81). In some embodiments, the cytidine deaminase is a TDD that comprises an amino acid sequence at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 92%, 94%, 95%, 96%, 97%, 98%, or 99% identical to the amino acid sequence of SEQ ID NO: 49 or 81. In certain embodiments, the first DddA portion comprises amino acids 1264-1333, 1264-1397, 1264-1404, 1264-1407, or a fragment thereof, of amino acids 1264-1427 of SEQ ID NO: 72; and the second DddA portion comprises the remainder, or a fragment thereof, of said amino acids of SEQ ID NO: 72; or vice versa; wherein the two portions form a functional cytidine deaminase. In certain embodiments, the first DddA portion comprises amino acids 1290-1333, 1290-1397, 1290-1404, 1290-1407, or a fragment thereof, of amino acids 1290-1427 of SEQ ID NO: 72; and the second DddA portion comprises the remainder, or a fragment thereof, of said amino acids of SEQ ID NO: 72; or vice versa; wherein the two portions form a functional cytidine deaminase. In some embodiments, the first and second DddA portions respectively comprise SEQ ID NOs: 82 and 83, SEQ ID NOs: 84 and 85, SEQ ID NOs: 18 and 19, SEQ ID NOs: 51 and 52, or SEQ ID NOs: 53 and 54; or vice versa.

[0014] In some embodiments of the base editor systems described herein, the cytidine deaminase is DddA that has a mutation at one or more residues selected from Y1307, T1311, S1331, V1346, H1366, N1367, N1368, P1369, E1370, G1371, T1372, F1375, V1392, P1394, P1395, I1399, P1400, V1401, K1402, A1405, and T1406 in SEQ ID NO: 72.

[0015] In some embodiments of the base editor systems described herein, the cytidine deaminase is a TDD that comprises the amino acid sequence of any one of SEQ ID NOs: 86-91 and 117-129. In certain embodiments, the cytidine deaminase comprises the toxic domain of a TDD comprising the amino acid sequence of any one of SEQ ID NOs: 86-91 and 117-129. In certain embodiments, the TDD comprises an amino acid sequence at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 92%, 94%, 95%, 96%, 97%, 98%, or 99% identical to the amino acid sequence of SEQ ID NO: 92, 95, 98, 101, 104, 107, 134, 143, 152, 157, 162, 167, 172, 177, 184, 189, 194, 199, 204, 209, 214, or 219. In particular embodiments, the cytidine deaminase is a TDD that comprises the amino acid sequence of SEQ ID NO: 92, 95, 98, 101, 104, 107, 134, 143, 152, 157, 162, 167, 172, 177, 184, 189, 194, 199, 204, 209, 214, or 219. In particular embodiments, the first and second cytidine deaminase portions respectively comprise SEQ ID NOs: 93 and 94, SEQ ID NOs: 96 and 97, SEQ ID NOs: 99 and 100, SEQ ID NOs: 102 and 103, SEQ ID NOs: 105 and 106, SEQ ID NOs: 108 and 109, SEQ ID NOs: 130 and 131, SEQ ID NOs: 132 and 133, SEQ ID NOs: 135 and 136, SEQ ID NOs: 137 and

138, SEQ ID NOs: 139 and 140, SEQ ID NOs: 141 and 142, SEQ ID NOs: 144 and 145, SEQ ID NOs: 146 and 147, SEQ ID NOs: 148 and 149, SEQ ID NOs: 150 and 151, SEQ ID NOs: 153 and 154, SEQ ID NOs: 155 and 156, SEQ ID NOs: 158 and 159, SEQ ID NOs: 160 and 161, SEQ ID NOs: 163 and 164, SEQ ID NOs: 165 and 166, SEQ ID NOs: 168 and 169, SEQ ID NOs: 170 and 171, SEQ ID NOs: 173 and 174, SEQ ID NOs: 175 and 176, SEQ ID NOs: 178 and 179, SEQ ID NOs: 180 and 181, SEQ ID NOs: 182 and 183, SEQ ID NOs: 185 and 186, SEQ ID NOs: 187 and 188, SEQ ID NOs: 190 and 191, SEQ ID NOs: 192 and 193, SEQ ID NOs: 195 and 196, SEQ ID NOs: 197 and 198, SEQ ID NOs: 200 and 201, SEQ ID NOs: 202 and 203, SEQ ID NOs: 205 and 206, SEQ ID NOs: 207 and 208, SEQ ID NOs: 210 and 211, SEQ ID NOs: 212 and 213, SEQ ID NOs: 215 and 216, SEQ ID NOs: 217 and 218, SEQ ID NOs: 220 and 221, or SEQ ID NOs: 222 and 223; or vice versa.

[0016] In a related aspect, the present disclosure also provides a fusion protein comprising i) a zinc finger protein (ZFP) domain that binds to gene (which may be a eukaryotic, e.g., human, gene) and ii) a cytidine deaminase polypeptide or a fragment thereof, e.g., wherein the cytidine deaminase is a TDD comprising an amino acid sequence at least 90% identical to SEQ ID NO: 49, 81, 92, 95, 98, 101, 104, 107, 134, 143, 152, 157, 162, 167, 172, 177, 184, 189, 194, 199, 204, 209, 214, or 219, optionally wherein the ZFP domain and the cytidine deaminase or fragment thereof are linked by a peptide linker. In some embodiments, the TDD comprises the amino acid sequence of SEQ ID NO: 49, 81, 92, 95, 98, 101, 104, 107, 134, 143, 152, 157, 162, 167, 172, 177, 184, 189, 194, 199, 204, 209, 214, or 219.

[0017] In a related aspect, the present disclosure provides a fusion protein comprising i) a zinc finger protein (ZFP) domain that binds to a gene (which may be a eukaryotic, e.g., human, gene), and ii) a cytidine deaminase inhibitory domain, e.g., wherein the cytidine deaminase is a TDD comprising an amino acid sequence at least 90% identical to SEQ ID NO: 49, 81, 92, 95, 98, 101, 104, 107, 134, 143, 152, 157, 162, 167, 172, 177, 184, 189, 194, 199, 204, 209, 214, or 219, optionally wherein the ZFP domain and the inhibitory domain are linked by a peptide linker. In some embodiments, the cytidine deaminase inhibitory domain is a TDDI, such as DddI where the cytidine deaminase is DddA. In some embodiments, the TDD comprises the amino acid sequence of SEQ ID NO: 49, 81, 92, 95, 98, 101, 104, 107, 134, 143, 152, 157, 162, 167, 172, 177, 184, 189, 194, 199, 204, 209, 214, or 219.

[0018] In a related aspect, the present disclosure provides a fusion protein comprising i) a zinc finger protein (ZFP) domain that binds to a gene (which may be a eukaryotic, e.g., human, gene), and ii) a nickase or a fragment thereof, optionally wherein the ZFP domain and the nickase or fragment thereof are linked by a peptide linker.

[0019] In one aspect, the present disclosure provides a pair of fusion proteins comprising a) a first fusion protein that comprises i) a zinc finger protein (ZFP) domain that binds to a gene (which may be a eukaryotic, e.g., human, gene), and ii) a first dimerization domain, and b) a second fusion protein that comprises i) a cytidine deaminase inhibitory domain, e.g., wherein the cytidine deaminase is a TDD comprising an amino acid sequence at least 90% identical to SEQ ID NO: 49, 81, 92, 95, 98, 101, 104, 107, 134, 143, 152, 157, 162, 167, 172, 177, 184, 189, 194, 199, 204, 209, 214, or 219, and ii) a second dimerization domain, wherein the first and second dimerization domains can dimerize in the presence of a dimerization-inducing agent. In some embodiments, the cytidine deaminase inhibitory domain is a TDDI, such as DddI where the cytidine deaminase is DddA. In some embodiments, the TDD comprises the amino acid sequence of SEQ ID NO: 49, 81, 92, 95, 98, 101, 104, 107, 134, 143, 152, 157, 162, 167, 172, 177, 184, 189, 194, 199, 204, 209, 214, or 219.

[0020] In another aspect, the present disclosure provides a pair of fusion proteins comprising a) a first fusion protein that comprises i) a zinc finger protein (ZFP) domain that binds to a gene (which may be a eukaryotic, e.g., human, gene), and ii) a first dimerization domain, and b) a second fusion protein that comprises i) a cytidine deaminase inhibitory domain, e.g., wherein the cytidine deaminase is a TDD comprising an amino acid sequence at least 90% identical to SEQ ID NO: 49, 81, 92, 95, 98, 101, 104, 107, 134, 143, 152, 157, 162, 167, 172, 177, 184, 189, 194, 199, 204, 209, 214, or 219, and ii) a second dimerization domain, wherein the first and second dimerization domains can dimerize in the absence of a dimerization-inhibiting agent. In some embodiments, the cytidine deaminase inhibitory domain is a TDDI, such as DddI where the cytidine deaminase is DddA. In some embodiments, the TDD comprises the amino acid sequence of SEQ ID NO: 49, 81, 92, 95, 98, 101, 104, 107, 134, 143, 152, 157, 162, 167, 172, 177, 184, 189, 194, 199, 204, 209, 214, or 219.

[0021] In one aspect, the present disclosure provides one or more nucleic acid molecules encoding the fusion protein(s) described herein, as well as expression constructs comprising the nucleic acid molecule(s) and viral vectors comprising the expression construct(s), optionally wherein the viral vectors may be an adeno-associated viral vector, an adenoviral vector, or a lentiviral vector. Also provided is a cell (which may be a eukaryotic cell, e.g., a mammalian cell or a plant cell) comprising a base editor system as described herein, fusion protein(s) as described herein, isolated nucleic acid molecule(s) as described herein, expression construct(s) as described herein, or viral vector(s) as described herein. In some

embodiments, the mammalian cell is a human cell, such as a human embryonic stem or a human induced pluripotent stem cell.

[0022] In some aspects, the present disclosure provides a method of changing a cytosine to a thymine in a target genomic region in a cell (which may be a eukaryotic cell, e.g., a mammalian or plant cell), comprising delivering a base editor system as described herein to the cell. In some embodiments, the change of the cytosine to the thymine creates a stop codon in the target genomic region. A multiplex format of the system may target more than one genomic region (e.g., 2, 3, 4, or 5 genomic regions). The editing may be performed *in vivo*, *ex vivo*, or *in vitro*.

[0023] Also provided are genetically engineered cells (which may be eukaryotic cells, e.g., mammalian cells such as human iPSCs or plant cells) obtained by the present editing methods.

[0024] Engineered cells described herein (e.g., engineered human cells), including pharmaceutical compositions comprising the cells and a pharmaceutically acceptable carrier, may be used for treating a patient in need thereof (e.g., a human patient in need thereof) or used in the manufacture of a medicament for treating a patient in need thereof. In some embodiments, the patient has cancer, an autoimmune disorder, an autosomal dominant disease, or a mitochondrial disorder. In some embodiments, the patient has sickle cell disease, hemophilia, cystic fibrosis, phenylketonuria, Tay-Sachs, prion disease, color blindness, a lysosomal storage disease, Friedreich's ataxia, or prostate cancer. Kits and articles of manufacture comprising the cells are also contemplated.

[0025] Other features, objects, and advantages of the invention are apparent in the detailed description that follows. It should be understood, however, that the detailed description, while indicating embodiments and aspects of the invention, is given by way of illustration only, not limitation. Various changes and modification within the scope of the invention will become apparent to those skilled in the art from the detailed description.

BRIEF DESCRIPTION OF THE FIGURES

[0026] **FIG. 1** is a schematic illustrating a pair of ZFP-TDD fusion proteins for C to T base editing. The rectangles represent DNA-binding zinc fingers in the ZFP domains of the fusion proteins. The arrow shapes above the underlined C nucleotide represent dimerized TDD domains of the fusion proteins. The black lines between the zinc finger domains and the TDD domains represent peptide linkers.

[0027] **FIG. 2A** is a schematic showing ZFP designs for *CCR5*-targeting ZFP-TDD fusion protein pairs. C9, C10, C18, and C24 are target nucleotides for base editing. Top strand (left to right): SEQ ID NO: 227. Bottom strand (right to left): SEQ ID NO: 228.

[0028] **FIG. 2B** is a schematic showing an example of a construct design for a dimerized ZFP-DddA pair. FLAG: FLAG tag. NLS: nuclear localization sequence. UGI: uracil DNA glycosylase inhibitor.

[0029] **FIG. 3** is a table showing the heatmap results of C to T base editing at a human *CCR5* locus by a series of ZFP-DddA fusion protein pairs. The degree of editing activity corresponds to the darkness of shading within a cell. L0, L7A, and L26 represent peptide linkers used to fuse the DddA domain to the C-terminus of the ZFP domain in the fusion protein.

[0030] **FIG. 4** is a table showing the heatmap results of C to T base editing at a human *CCR5* locus by a series of ZFP-DddA fusion protein pairs, wherein the DddA split occurs at different positions. The degree of editing activity corresponds to the darkness of shading within a cell.

[0031] **FIG. 5** is a schematic showing ZFP designs for *CCR5*-targeting ZFP-TDD fusion proteins. C9, C10, C18, and C24 are target nucleotides for base editing. From top to bottom: SEQ ID NO: 229 (left to right), SEQ ID NO: 230 (right to left), SEQ ID NO: 231 (left to right), SEQ ID NO: 232 (right to left), SEQ ID NO: 233 (left to right), and SEQ ID NO: 234 (right to left).

[0032] **FIGS. 6A-6C** are tables showing the heatmap results of C to T base editing at a human *CCR5* locus by a series of ZFP-DddA fusion protein pairs with the indicated DddA mutations. The mutations are numbered with respect to SEQ ID NO: 72. The degree of editing activity corresponds to the darkness of shading within a cell

[0033] **FIG. 7A** is a schematic illustrating the combined use of the ZFP-TDD base editing system and a nickase system for increasing base editing efficiency. The nickase system shown here is a CRISPR/Cas-based nickase system. The illustrative gene locus is a human *CCR5* locus. Top strand (left to right): SEQ ID NO: 235. Bottom strand (right to left): SEQ ID NO: 236.

[0034] **FIG. 7B** is a table showing the heatmap results of DddA C to T base editing at a human *CCR5* locus using the approach of **FIG. 7A**. The degree of editing activity corresponds to the darkness of shading within a cell.

[0035] **FIG. 8** is a schematic illustrating the combined use of the ZFP-TDD base editing system and a CRISPR/Cas-based nickase system.

[0036] **FIG. 9** is a schematic illustrating an example of a trimeric ZFP-TDD + FokI nickase base editing system.

[0037] **FIG. 10** is a schematic showing ZFP designs for combined use of *CCR5*-targeting ZFP-TDD fusion protein pairs with a ZFP-nickase. C9, C10, C18, and C24 are target nucleotides for base editing. Top strand (left to right): SEQ ID NO: 237. Bottom strand (right to left): SEQ ID NO: 238.

[0038] **FIG. 11** is a table showing the heatmap results of DddA C to T base editing at a human *CCR5* locus using the approach of **FIG. 10**. The degree of editing activity corresponds to the darkness of shading within a cell.

[0039] **FIG. 12** is a table showing the heatmap results of C to T base editing at a human *CCR5* locus by a series of ZFP-TDD fusion protein pairs. The degree of editing activity corresponds to the darkness of shading within a cell. O1: TDD1; O2: TDD2; O3: TDD3; O4: TDD4; O5: TDD5; O6: TDD6.

[0040] **FIG. 13** is a table showing the heatmap results of the highest frequency of C to T base editing for any C in the *CCR5* base editing window by ZFP fusion protein pairs with TDD1-TDD6. O1: TDD1; O2: TDD2; O3: TDD3; O4: TDD4; O5: TDD5; O6: TDD6.

[0041] **FIG. 14** is a table showing the heatmap results of the highest frequency of C to T base editing for any C in the *CCR5* base editing window by ZFP fusion protein pairs with TDD1-TDD6. O1: TDD1; O2: TDD2; O3: TDD3; O4: TDD4; O5: TDD5; O6: TDD6.

[0042] **FIG. 15** is a schematic showing ZFP designs for *CIITA*-targeting ZFP-TDD fusion protein pairs. G2, G5, C6, C8, G10, G11, G14, C15, and C16 are target nucleotides for base editing. Top strand (left to right): SEQ ID NO: 239. Bottom strand (right to left): SEQ ID NO: 240.

[0043] **FIG. 16** is a table showing the heatmap results of the highest frequency of C to T base editing at a human *CIITA* locus ("site 2") by a series of ZFP-TDD fusion protein pairs. The degree of editing activity corresponds to the darkness of shading within a cell. O1: TDD1; O14: TDD14; etc.

[0044] **FIG. 17** is a table showing the heatmap results of the highest frequency of C to T base editing for any C (underlined) in the *CIITA* base editing window and its sequence motif for DddA, TDD4, TDD6, TDD9, TDD10, TDD14, TDD15 and TDD18. Amplicon: SEQ ID NO: 244. O4: TDD4; O6: TDD6; etc.

[0045] **FIG. 18** is a table showing the heatmap results of C to T base editing at a human *CIITA* locus ("site 2") by a ZFP fusion protein pair with TDD6 or TDD14. L26, L21, L18, L13, L11, L9, L6, and L4 represent peptide linkers used to fuse the TDD6 or TDD14 domain

to the C-terminus of the ZFP domain in the fusion protein. The degree of editing activity corresponds to the darkness of shading within a cell. O6: TDD6; O14: TDD14.

[0046] **FIG. 19** is a schematic illustrating a design for inhibition of a TDD with a targeted ZFP-TDDI.

DETAILED DESCRIPTION OF THE INVENTION

[0047] The present disclosure provides systems and methods for base editing, e.g., from cytosine (C) to thymine (T), in cellular DNA such as genomic DNA. The systems entail the use of ZFP-toxin-derived deaminase (TDD) fusion proteins (ZFP-TDDs). By providing precise gene editing in a cellular context, the present systems and methods can be used for the prevention and/or treatment of numerous diseases. It is contemplated that these systems and methods will be particularly useful for cell-based therapies that require the simultaneous knock-out of multiple human genes.

[0048] The present systems and methods can convert targeted C:G base pairs to T:A base pairs. In some embodiments, the base editing systems may also include proteins (e.g., UGI) that increase the stability of the conversion, and/or endonucleases that nick the DNA near the targeted base so as to stimulate DNA repair in the edited region and to promote the correction of the G nucleotide on the opposite strand to A, forming the edited T:A base pair.

[0049] The present systems and methods are advantageous in part due to the compact size of the ZFP domains in the fusion proteins. In comparison, the large physical size of a TALE and the long C-terminal TALE linker may limit how small the base editing window can be, as well as design density. The size and highly repetitive nature of engineered TALEs also make it challenging to deliver TALE-based base editors to human cells using common viral vectors. The present ZFP-derived base editing systems circumvent these problems. For instance, the compactness of these ZFP-derived systems may allow for packaging within a single AAV vector, in contrast to TALE base editor systems (e.g., TALE-TDDs) or CRISPR/Cas base editor systems. In addition, due to the small size of the fusion proteins herein, it is possible to include a nickase in the editing system so as to allow the generation of a DNA nick near the edited base and thereby facilitate the DNA repair machinery to change the base opposite the edited C from G to a corresponding A, forming the correct T:A base pair. The inclusion of a nickase may greatly increase the base editing efficiency.

I. Zinc-Finger Fusion Proteins

[0050] Provided are fusion proteins that contain a DNA-binding zinc finger protein (ZFP) domain fused to a base editor domain (e.g., a cytidine deaminase domain, which may be a TDD such as one described herein), a cytidine deaminase inhibitor (e.g., a TDDI, such as DddI where the cytidine deaminase is DddA) domain, and/or a nickase domain (e.g., a FokI domain). As used herein, a "fusion protein" refers to a polypeptide where heterologous functional domains (i.e., functional domains that are not naturally present in the same protein in nature) are covalently linked (e.g., through peptidyl bonds). These fusion proteins, which can be recombinantly made, are components of the present base editor systems. In some embodiments, a ZFP fusion protein herein comprises a cytidine deaminase domain (e.g., derived from a TDD as described herein) and additionally a nickase domain and/or a UGI domain.

[0051] Other formats of the present systems also are contemplated herein. For example, instead of peptidyl links, two functional domains may be brought together by noncovalent bonds. In some embodiments, two functional domains (e.g., a ZFP domain and a cytidine deaminase inhibitor domain; or a ZFP domain and a nickase domain) each are fused to a dimerization partner (e.g., leucine zipper and those described further herein), such that the two functional domains are brought together through interaction of the dimerization partners. In certain embodiments, the dimerization of these domains may be controlled by the presence or absence of a specific agent (e.g., a small molecule or peptide). It is contemplated that such formats may substitute for fusion proteins in any aspect of the present invention.

[0052] Each component of the present base editor systems is further described in detail below.

A. Base Editors

[0053] The ZFP-cytidine deaminase fusion proteins of the present disclosure comprise a cytidine deaminase domain in addition to a ZFP domain. The term "deaminase" or "deaminase domain," as used herein, refers to a protein that catalyzes a deamination reaction. A cytidine deaminase domain, for example, may catalyze the deamination of cytosine to uracil, wherein the uracil is replaced by a thymine base during DNA replication or repair. The deaminase domain may be naturally-occurring or may be engineered. In some embodiments, a cytidine deaminase of the present disclosure operates on double-stranded DNA.

[0054] In some embodiments, the cytidine deaminase is derived from a toxin that may be, e.g., from a prokaryotic or eukaryotic organism. In certain embodiments, the organism may

be bacteria or fungus. Such a cytidine deaminase is referred to herein as a toxin-derived deaminase (TDD). DddA and DddA orthologs are TDDs. As used herein, a cytidine deaminase "derived from" a toxin may refer to a cytidine deaminase that is the same as the naturally occurring toxin or is a modified version of the toxin that retains deaminase activity.

[0055] In some embodiments, the cytidine deaminase is DddA (SEQ ID NO: 72). In certain embodiments, the cytidine deaminase comprises the toxic domain (e.g., amino acids 1290-1427 (SEQ ID NO: 49) or 1264-1427 (SEQ ID NO: 81)) of DddA, and the fusion protein is termed ZFP-DddA. An exemplary full sequence of the DddA protein derived from *Burkholderia cenocepacia* is shown below:

```
        10         20         30         40         50
MYEAARVTDP IDHTSALAGF LVGAVLGIAL IAAVAFATFT CGFGVALLAG
        60         70         80         90        100
MMAGIGAQAL LSIGESIGKM FSSQSGNIIT GSPDVYVNSL SAAYATLSGV
       110        120        130        140        150
ACSKHNPIPL VAQGSTNIFI NGRPAARKDD KITCGATIGD GSHDTFFHGG
       160        170        180        190        200
TQTYLPVDDE VPPWLRTATD WAFTLAGLVG GLGGLLKASG GLSRAVLPCA
       210        220        230        240        250
AKFIGGYVLG EAFGRYVAGP AINKAIGGLF GNPIDVTTGR KILLAESETD
       260        270        280        290        300
YVIPSPLPVA IKRFYSSGID YAGTLGRGWV LPWEIRLHAR DGRLWYTDAQ
       310        320        330        340        350
GRESGFPMLR AGQAAFSEAD QRYLTRTPDG RYILHDLGER YYDFGQYDPE
       360        370        380        390        400
SGRIAWVRRV EDQAGQWYQF ERDSRGRVTE ILTCGGLRAV LDYETVFGRL
       410        420        430        440        450
GTVTLVHEDE RRLAVTYGYD ENGQLASVTD ANGAGVRQFA YTNGLMTNHM
       460        470        480        490        500
NALGFTSSYV WSKIEGEPRV VETHTSEGEN WTFEYDVAGR QTRVRHADGR
       510        520        530        540        550
TAHWRFDAQS QIVEYTDLDG AFYRIKYDAV GMPVMLMLPG DRTVMFEYDD
       560        570        580        590        600
AGRIIAETDP LGRTTRTRYD GNSLRPVEVV GPDGGAWRVE YDQQGRVVSN
       610        620        630        640        650
QDSLGRENRY EYPKALTALP SAHIDALGGR KTLEWNSLGK LVGYTDCSGK
       660        670        680        690        700
```

TTRTSFDAFG RICSRENALG QRITYDVRPT GEPRRVTYPD GSSETFEYDA
710 720 730 740 750
AGTLVRYIGL GGRVQELLRN ARGQLIEAVD PAGRRVQYRY DVEGRLRELQ
760 770 780 790 800
QDHARYTFTY SAGGRLLTET RPDGILRRFE YGEAGELLGL DIVGAPDPHA
810 820 830 840 850
TGNRSVRTIR FERDRMGVLK VQRTPTEVTR YQHDKGDRLV KVERVPTPSG
860 870 880 890 900
IALGIVPDAV EFEYDKGGRL VAEHGSNGSV IYTLDELDNV VSLGLPHDQT
910 920 930 940 950
LQMLRYGSGH VHQIRFGDQV VADFERDDLH REVSRTQGRL TQRSGYDPLG
960 970 980 990 1000
RKVWQSAGID PEMLGRGSGQ LWRNYGYDAA GDLIETSDSL RGSTRFSYDP
1010 1020 1030 1040 1050
AGRLISRANP LDRKFEEFAW DAAGNLLDDA QRKSRGYVEG NRLLMWQDLR
1060 1070 1080 1090 1100
FEYDPFGNLA TKRRGANQTQ RFTYDGQDRL ITVHTQDVRG VVETRFAYDP
1110 1120 1130 1140 1150
LGRRIAKTDT AFDLRGMKLR AETKRFVWEG LRLVQEVRET GVSSYVYSPD
1160 1170 1180 1190 1200
APYSPVARAD TVMAEALAAT VIDSAKRAAR IFHFHTDPVG ALQEVTDEAG
1210 1220 1230 1240 1250
EVAWAGQYAA WGKVEATNRG VTAARTDQPL RFAGQYADDS TGLHYNTFRF
1260 1270 1280 1290 1300
YDPDVGRFIN QDPIGLNGGA NVYHYAPNPV GWVDPWGLAG SYALGPYQIS
1310 1320 1330 1340 1350
APQLPAYNGQ TVGTFYYVND AGGLESKVFS SGGPTPYPNY ANAGHVEGQS
1360 1370 1380 1390 1400
ALFMRDNGIS EGLVFHNNPE GTCGFCVNMT ETLLPENAKM TVVPPEGAIP
1410 1420
VKRGATGETK VFTGNSNSPK SPTKGGC (SEQ ID NO: 72)

As used herein, unless specified otherwise, the term "DddA" refers to the DddA toxic domain.

[0056] In certain embodiments, the cytidine deaminase is a "re-wired" version of DddA (e.g., SEQ ID NO: 50).

[0057] The present disclosure also provides variants of DddA mutated at residues that form the nucleotide pocket (e.g., Y1307, T1311, S1331, V1346, H1366, N1367, N1368,

P1369, E1370, G1371, T1372, F1375, V1392, P1394, P1395, I1399, P1400, V1401, K1402, A1405, T1406, or any combination thereof, wherein the numbering of the residues is with respect to SEQ ID NO: 72). The DddA may be mutated, for example, at 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, or 22 of said residues. In some embodiments, DddA is mutated at residue E1370, N1368, Y1307, T1311, S1331, K1402, or any combination thereof. In certain embodiments, DddA is mutated at residue E1370, N1368, Y1307, or any combination thereof. In certain embodiments, the mutation(s) may increase DddA efficiency, increase DddA activity, change the DddA activity window, or any combination thereof. It is contemplated that such variants may substitute for wild-type DddA in any aspect of the present invention.

[0058] In particular embodiments, the cytidine deaminase domain (e.g., derived from a TDD described herein) is a "split enzyme" comprised of first and second "half domains" or "splits" that lack cytidine deaminase activity alone but dimerize to form an active cytidine deaminase. As used herein, half domains that are "inactive" or "lack cytidine deaminase activity" may be half domains that i) lack any cytidine deaminase activity (e.g., any detectable cytidine deaminase activity), ii) lack specific cytidine deaminase activity, or iii) lack significant cytidine deaminase activity (i.e., on-target base editing activity of 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, or 10% or more, which in particular embodiments may be 10% or more). For example, assembly of the active cytidine deaminase may be driven by the binding of half domain-linked zinc finger proteins to DNA targets in proximity to each other such that the half domains are positioned to allow assembly of a functional cytidine deaminase.

[0059] It is understood that the "half domain" pairs described herein may refer to any pair of cytidine deaminase polypeptide sequences that separately lack cytidine deaminase activity, but together form a functional cytidine deaminase domain (either wild-type or a variant discussed herein). Where the cytidine deaminase is DddA, the "split" in the DddA sequence may occur at any of a number of positions, such as, for example, at G1322, G1333, A1343, N1357, G1371, N1387, E1396, G1397, A1398, I1399, P1400, V1401, K1402, R1403, G1404, A1405, T1406, G1407, or E1408, and need not be in the middle of the protein. In some embodiments, the "split" occurs at G1322, G1333, A1343, N1357, G1371, N1387, G1397, G1404, or G1407. In certain embodiments, the "split" occurs at G1404, G1407, G1333, or G1397. In particular embodiments, the "split" occurs at G1404 or G1407. In some embodiments, the DddA half domain pairs may comprise the amino acid sequences of:

a) SEQ ID NOs: 82 and 83;

b) SEQ ID NOs: 84 and 85;

c) SEQ ID NOs: 18 and 19;

d) SEQ ID NOs: 51 and 52; or

e) SEQ ID NOs: 53 and 54.

[0060] In certain embodiments, the TDD may comprise, for example, an amino acid sequence under NCBI Accession No. WP_069977532.1 ("TDD1," SEQ ID NO: 86), WP_021798742.1 ("TDD2," SEQ ID NO: 87), QNM04114 ("TDD3," SEQ ID NO: 88), WP_181981612 ("TDD4," SEQ ID NO: 89), AXI73669.1 ("TDD5," SEQ ID NO: 90), WP_195441564 ("TDD6," SEQ ID NO: 91), AVT32940.1 ("TDD7," SEQ ID NO: 117), WP_189594293.1 ("TDD8," SEQ ID NO: 118), TCP42004.1 ("TDD9," SEQ ID NO: 119), WP_171906854.1 ("TDD10," SEQ ID NO: 120), WP_174422267.1 ("TDD11," SEQ ID NO: 121), WP_059728184.1 ("TDD12," SEQ ID NO: 122), WP_133186147.1 ("TDD13," SEQ ID NO: 123), WP_083941146.1 ("TDD14," SEQ ID NO: 124), WP_082507154.1 ("TDD15," SEQ ID NO: 125), WP_044236021.1 ("TDD16," SEQ ID NO: 126), WP_165374601.1 ("TDD17," SEQ ID NO: 127), NLI59004.1 ("TDD18," SEQ ID NO: 128), or KAB8140648.1 ("TDD19," SEQ ID NO: 129), or a part of said amino acid sequence that is capable of cytidine deaminase activity (e.g., a "toxic domain"). These amino acid sequences are shown below:

NCBI Accession No. WP_069977532.1 (TDD1)

MSSSDAGRAFGVPENVLARFTRYPGGARRRAGRTARARRLGIVLSAVLSATLLPAEAWAIAP
PAPRTGPTLDALQQEEEVDPDPAAMEELDDWDGGPVEPPADYTPTEVTPPTGGTAPVPLDSA
GEELVPAGTLPVRIGQASPTEEDPAPPAPSGTWDVTVEPRATTEAAAVDGAIIKLTPPASGS
TPVDVELDYGRFEDLFGTEWSSRLKLTQLPECFLTTPELEECGTPITIPTSNDPATGTVRAT
VDPADGQPQGLAAQSGGGPAVLAATDSASGAGGTYKATSLSATGSWTAGGSGGGFSWSYPLT
IPDTPAGPAPKISLSYSSQSVDGRTSVANGQASWIGDGWDYHPGFVERRYRSCNDDRSGTPN
NDNSADKEKSDLCWASDNVVMSLGGSTTELVRDDTTGTWVAQNDTGARIEYKDKDGGALAAQ
TAGYDGEHWVVTTRDGTRYWFGRNTLPGRGAPTNSALTVPVFGNHTGEPCHAATYAASSCTQ
AWRWNLDYVEDVHGNAMVVDWKKEQNRYAKNEKFKAAVSYDRDAYPTQILYGLRADDLAGPP
AGKVVFHAAPRCLESAATCSEAKFESKNYADKQPWWDTPATLHCKAGDENCYVTSPTFWSRV
RLSAIETQGQRTPGSTALSTVDRWTLHQSFPKQRTDTHPPLWLESITRVGFGRPDASGNQSS
KALPAVTFLPNKVDMPNRVLKSTTDQTPDFDRLRVEVIRTETGGETHVTYSAPCPVGGTRPT
PASNGTRCFPVHWSPDPAAFSDENLDKSGYEPPLEWFNKYVVTKVTEMDLVAEQPSVETVYT
YEGDAAWAKNTDEYGKPALRTYDQWRGYASVVTRTGTTANTGAADATEQSQTRTRYFRGMSG
DAGRAKVHVTLTDVTGTATTVEDLLPYQGMAAETLTYTKAGGDVAARELAFPYSRKTASRAR
PGLPALEAYRTGTTRTDSIQHISGDRTRAAQNHTTYDDAYGLPTQTYSLTLSPNDSGTLVAG
DERCTVTTYVHNTAAHIIGLPDRVRATTGDCAAAPNATTGQIVSDSRTAYDALGAFGTAPVK
GLPVQVDTISGGGTSWITSARTEYDALGRATKVTDAAGNSTTTTYSPATGPAFEVTVTNAAG
HATTTTLDPGRGSALTVTDQNGRKTTSTYDELGRATGVWTPSRPVNQDASVRFVYQIEDSKV
PAVHTRVLRDAGTYEESIELYDGFLRPRQTQREALGGGRIVTETLYNANGSAKEVRDGYLAE
GEPARELFVPLSLDQVPSATRTAYDGLGRPVRTTTLHRGVPRHSATTAYGGDWELSRTGMSP
DGTTPLSGSRAVKATTDALGRPARIQHFTTQNVSAESVDTTYTYDPRGPLAQVTDAQQNTWT
YTYDARGRKTSSTDPDAGAAYFGYNALDQQVWSKDNQGRLQYTTYDVLGRQTELRDDSASGP

LVAKWTFDTLPGAKGHPVASTRYNDGAAFTSEVTGYDTEYRPTGNKVTIPSTPMTTGLAGTY
TYASTYTPTGKVQSVDLPATPGGLAAEKVITRYDGEDSPTTMSGLAWYTADTFLGPYGEVLR
TASGEAPRRVWTTNVYDEDTRRLTRTTAHRETAPHPVSTTTYGYDTVGNITSIADQQPAGTE
EQCFSYDPMGRLVHAWTDGNSAVCPRTSTAPGAGPARADVSAGVDGGGYWHSYAFDAIGNRT
KLTVHDRTDAALDDTYTYTYGKTLPGNPQPVQPHTLTQVDAVLNEPGSRVEPRSTYAYDTSG
NTTQRVIGGDTQTLAWDRRNKLTSVDTNNDGTPDVKYLYDASGNRLVEDDGTTRTLFLGEAE
IVVNTAGQAVDARRYYSSPGAPTTIRTTGGKTTGHKLTVMLSDHHSTATTAVELTDTQPVTR
RRFDPYGNPRGTEPTTWPDRRTYLGVGIDDPATGLTHIGAREYDASTGRFISVDPVMDLTDP
LQMNGYTYANADPINNSDPTGLLLDARGGGTQKCVGTCVKDVTNRKGIPLPPGEEWKHEGEA
QTDFNGDGFITVFPTVNVPAKWKKAKKYTEAFYKAVDTACFYGRESCADPEYPSRAHSINNW
KGKACKAVGGKCPERLSWGEGPAFAGGFAIAAEEYAGRGGYRGGGARRGSPCKCFLAGTEVL
MADGSTKSIEDIKLGDEVVATDPVTGEAGAHPVSALIATENDKRFNELVIITSEGVERLTAT
HEHPFWSPSEGEWLEAGELRTGMTLRSDSGETLVVAGNRAFTQRARTYNLTVADLHTYYVLA
GQTPVLVHNANCGPHLKDLQKDYPRRTVGILDVGTDQLPMISGPGGQSGLLKNLPGRTKANG
EHVETHAAAFLRMNPGVRKAVLYIDYPTGTCGTCRSTLPDMLPEGVQLWVISPRRTEKFTGL
PD (SEQ ID NO: 86)

NCBI Accession No. WP_021798742.1 (TDD2)

MVDLGAYEEPVAFDDGVADALRSAASALSGTLSGQAASRSSWAATASTDFEGHYADVFDANA
RAACDDCSNIASALDALAADVQTMKDAAASERDRRRQAKEWADRQKDEWAPKSWIDDHLGLD
KPPAGPPETPVVDAQAPTVATWSEPAQGQAGGVSSARPDDLRTYSSNVTGANDTVTTQKGTL
DGALSDFADRCSWCSIDTSGITTALAAFGANNTNETRWVDTVAAAFEAAGGSGAISAVSDAA
LDASLQAAGVTQSRQPVDVTAPTIQGDPQTSGYADDPVNTTTGNFIEPETDLAFSGGCASLG
FDRVYNSLSAGVGAFGPGWASTADQRLLVTEDGAVWVQPSGRHVVFPRLGNGWDRAHNDTYW
LHTTTDTTGPTPGDAPTTGAAGGAGVFVVSDNAGGRWVFDRAGRPVSVSRGPGTRVDHRWDG
DRLVGLTHERGRAVTIEWNDHHTRITALTANDGRRVDYGYDPAGRLTEAASAGGTRTYGWNE
AGLIATVTDPDGVVEAANTYNEHGQVTSQRSRFGRLSHYTYLPGGVTQVADEDGGRANTWIH
DQTGRLVGMVDADGNRQSIGWDQWGNRVQITGRDGRTTVCRYDARGRLITRQEASGARTDYE
WDEADRVVQVTVTDTTSSSHGNTSSAGGSGPSVTSYEYEGAGRNPSTVTDPEGGVTRLTWDQ
NLLTEATDPAGVRVRLGYDGHGDLVSTTNAAGDTARLVRDGAGRVVAAITPLGHRTEYRYDE
AGRLASRQDPDGALWRYEHTTGGRLSAVVDPDGGRTVTEYGPGGVEEATTDPLGRRLEQEWD
DLGNLAGVRLPDGREWSYVHDGLSRLTETVDPAGGLWRREYDVNGMVAATVDPTGVRRGLAW
AADGSVTVSDASGTARVGVDGLGRPVSVSVSSAPAPGEAVPMGMSLEETVGTGAPAPGGAGP
DGPDARVVVRDLCGRPVEALDADGGLTRLMRDAAGRLVEEISPAGRSTRYEWDRCGRLSAVI
GPDGARTTMAYDAASRLIAQDGPGGRVRVAYDRCGRLSTVTAPGRGKTTWGYDRAGRVRSVR
SPAWGLVRFGYDPAGQLTAVTNALGGVTRYDYDECGRLVQVTDPLGHVTRRTYTAADRVETL
VDPLGRTTQAGYDAAGRQLWQTDDTGERLAFGWDEAGRLERVATGGEGLPGQTCCALTRPGR
RVLRVTGPGGARDELVFDRLGRLARHARGGRTVGEWSWDPDGACTAFTGPDGQRVRYAYDDA
GALVRVEGTAFGPVTVRRDTAGRLTGMDGPGLTQRWDRDETGHVIAYRRTKNGVTTSSRVSR
DESGRVTAVDGPDGTVRYGYDPAGQLARIEGPDGRRESFTWDKAGHLTRRSVERPGARPETT
LYSYDPAGQLASTDGPDGRTLYTWDAAGRRTGQDGPDGHWSYSWAPSGHLTAVTRRTPHDAR
TWRISRDGLGLPRRIDDTDLAWDLSGPVPALTRFGTHTVTGLPRALAIDGTLTSTGWRPARP
TSADDPWAPPPPVVETDGARLGVGGAVGLGGLEILGARVHDPTTFSFLSPDPLDQPPLAPWA
TNPYSYAANNPLAFTDPTGLRPLTDTDFEAYKHDHGGLGGWIADHKDYLIGGAMVIAGGVLM
ATGVGGPLGGMLIGAGADTIIQRATTGQVDYGQVAVSGLLGAAGGGAASALLKGGGRLATEL
GATGLRTAITTGAASGTASGAGGSGYGYLTGPGPHTVSGFLTSTATGAVEGGLLGGASGAAG
HGLSTTGKNVLGHFEPTPTTPQGTSSDTIAEMLNSASQPGRTAGVLDIDGELTPLTSGRPSL
PNYIASGHVEGQAAMIMRQQQVQSATVYHDNPNGTCGYCYSQLPTLLPEGAALDVVPPAGTV
PPSNRWHNGGPSFIGNSSEPKPWPR (SEQ ID NO: 87)

NCBI Accession No. QNM04114.1 (TDD3)

MSLPEYDGTTTHGVLVLDDGTQIGFTSGNGDPRYTNYRNNGHVEQKSALYMRENNISNATVY
HNNTNGTCGYCNTMTATFLPEGATLTVVPPENAVANNSRAIDYVKTYTGTSNDPKISPRYKG
N (SEQ ID NO: 88)

NCBI Accession No. WP_181981612 (TDD4)

MLAIEKIKSGDKVISTDPETMETSPKTVLETYIREVTTLVHLTVNGEEIVTTVDHPFYVKNQ
GFIKAGELIVGDELLDSNCNVLLVENHSVELTDEPVTVYNFQVEDFHTYHVGKCRLLVHNAN
CNQEKPVLPKYDGKTTEGVMVTPDGKQISFKSGNSSTPSYPQYKAQSASHVEGKAALYMREN
GINEATVFHNNPNGTCGFCDRQVPALLPKGAKLTVVPPSNSVANNVRAIPVPKTYIGNSTVP
KIK (SEQ ID NO: 89)

NCBI Accession No. AXI73669.1 (TDD5)

MSSSVSGRAFRVSGVLTRITKSWTPGSARRSSASVRHRGRAVRARSLGVTLSAVLAATLLPA
EAWAIAPPAPRIGPSLVDLQQEEPADPDQAKIDELSTWSGAPVEPPADYTPTATTPPAGGTA
PVALDGAGDDLVPVGNLPVRLGKASPTDEEPDPPAPGGTWDVAVEPRTSTEASDVDGALITV
TPPSGGATPVDIELDYGKFEDLFGTAWSSRLRLTQLPECFLTTPELDECTTVVDVPSVNDPS
NDTVRATIDPAASPQQGLSTQSGGGPVVLAATDSASGAGGTYKATPFTATGTWTAGGSGGGF
SWSYPLTAPAPPAGPAPTISLSYSSQSVDGRTSVANGQASWIGDGWDYNPGFIERRYRSCND
DRSGTPNNAGGKDKKKSDLCWASDNLVMSLGGSATALVHDGTTGAWVAQSDTGARIEYRTRT
GSPKTAQTGAYDGEYWVVTTRDGTRYWFGRNTIPGRTAATESALTVPVFGNHSGEPCHATAY
ADSSCAQAWRWNLDYVEDVHGNAMIVDWKKETNRYARNEKFKEAVAYHRGGYPAQILYGLRA
DDLNGAPAGKVVFKTAPRCVEDAGTTCSPTGYESDNYADKQPWWDTPATLHCKSGAKNCFVT
SPTFWSSVRLTEIETHGRRTPGSTALSLVDSWTLKQSFPKQRTDTHPPLWLESITRTGHGAP
NASGEQTSRALPPVSFLPNVVDMPNRVSKGATDETPDFDRLRVETVRTETGGEIHVDYSAPC
AVGTAHPSPETNTTRCFPVHYSPDPEALSDEVLAKKPAPVEWFNKYVVQKVTEKDRVARQPD
VVTTYAYEGGGAWGRSTDEFTKPKLRTYDQWRGYASVLVRKGVTGADPAAADATEQSQTRMR
YFRGMSGDAGRPTVTVKDSTGAETLGEDLAPYQGMPAETVAYTRAGGDVASRILAWPTSRET
ASQARPGLPALKAHRVATARTETVETISGGRTRTARTVTTYDDTYGLPLTAETLTLTPDGSG
GTTTGDRSCSTNTYVHNTAKHLIGLVQRARTTVGTCAQAATASGSDVVSDTRVSYDALDAFG
AAPVRGLPFRTDTVGADGTGWVTSARTEYDPLGRATEVRDAKGHVSKVGFVPPTGPAFTTTS
TDAKGHTTTTALDPARGTALSVTDANGRRTTSAYDELGRTTAVWSPSRTQGTDKASVLFDYQ
IEDNKVPATRTRVLRDNGTYEDSVTVYDGLLRPRQAQTEALGGGRIVTETLYNANGAPAETR
NGYLAEGEPQTELFVPLSLTQVPSASKTAYDGLGRAVRTTVLHAGDPQHSATVRHEGDRTLT
RTGMSADGTTPMPGSRSTATWTDALGRTSKIEHFTATDLSAAIDTRYTYDARGNLAKVTDAR
DNIWTYTYDARGRLTFSTDPDAGSSSFGYDVLDRQIWSKDSRQRSQHTVYDELGRRTELRDD
SAEGPLVAKWTYDTLPGAKGLPVASTRYHEGAEFTSEVTGYDQEYRPTGSRTTIPSTPLTTG
LAGTYTYKNTYTPTGLPQSVELPATPGGLAAEKVITRYDGEGSPRTTSGLAWYTVDTVLSPL
GQVLRTASGEAPNRVWATHFYDESTGRLDRRITDRETLDPSRISETSYAHDTVGNITSITDT
QSPARVDRQCFAYDPMGRLAHAWTAKSPGCPRSSTAQGAGPNRTDVSPSIDGAGYWHSYEFD
TIGNRTGMVVHDPADPALDDTYVYTHGVPSEGPLQPATLQPHTLTKVDATVRGPGSTVTSSS
TYAYDPSGNTTQRVIGGDTQALTWDRRNKLMSADTDDDGTADVTYLYDASGNRLLEADATTR
TLYLGESEIVVDTAGRPVEARRYYSHPGAPTTLRTTGGRTSGHTLTVQLTDHHNTPTASVAL
TGGQPVTRRMFDPYGNPRGTEPTTWPDRRTYLGVGIDDETTGLTHIGAREYDSVTGRFISAD
PIIDIADPLQMNGYAYANNNPVTNWDPTGLKSDECGSLYRCGGNQVITTKTTKYQDVNTVAR
HFEKTASWATLAQWKAEGLGKSPAFGKAKKLTKWKNEHYEKNWTINLVPGMARSWVSGVDAA
ASAIMPFPTVQAAPLYDSLVSSLGVNTKGRAYANGEGLMDGLSMVGGVGAIAPGIKSGLKAA
AKGCGPGNSFTPGTEVALADGTTKPIEDIKIGDEVLATDPETGETRAEKVTAEIRGDGTKNL
VKVTIDTDGDRGTDTAEITATDGHPFWVPELGRWIDATDLAPGQWLRTSAGTHVQITAIKRW
TETATVHNLTVADLHTYYVLAGKTPVLVHNENCGPNLKDLPKDYDRRTVGILDVGTDQLPMI
SGPGGQSGLLKNLPGRTKANTDHVEAHTAAFLRMNPGIRKAVLYIDYPTGTCGTCGSTLPDM
LPEGVQLWVISPRKTEKFAGLPD (SEQ ID NO: 90)

NCBI Accession No. WP_195441564 (TDD6)

MKLTYKELEIELELAGLLAVEELVLTQGLNCHAGLTLKILIEEEQRDELVTMSSDAGVTVRE
LEKTNGQVVFRGKLETVSARRENGLFYLYLEAWSYTMDWDRVKKSRSFQNGALTYMEVVQRV
LSGYGQSGVTDHATGGACIPEFLLQYEESDWVFLRRLASHFGTYLLADATDACGKVYFGVPE
ISYGTVLDRQGYTMEKDMLHYARVLEKEGVLSQEASCWNVTVRFFLRMWETLTFNGIEAVVT
AMRLHTEKGELVYSYVLARRAGIRREKEKNPGIFGMSIPATVMERSGNRIRVHFEIDPEYEA

SEKTKYFTYAIESSSFYCMPEEGSQVHIYFPDHDEQGAVAVHAIRSGEGASGSCSTPENKRF
SDPSGSAMDMTPASLQFAPDAGGATVLHLEGGGFLSLTGMDIKLKTQMGMASDKEKPMQDLM
ICGEQKLTMQIGESSDDCIVMEAGTEVRSALVVQEADSSPAAVPSGDELLSEQEAADAQARE
AENNAVKEDMITKKQESKRKIVDGVISLVTVVGLTALTVATGGLAAPFAIAAGVKATFAVAD
IAEGLDGYSKMNAMDASRPANFLRDTVFGGNQTAYDITSMITDVAFDVVSGKALVGAFSGAD
KVSKVQKFAGKAMSFWNGICPKTKVANFLFQMGGTMLFGAVNDYLTTGKVDLKNLGLDAFAG
LAKGTLGTAGTEKIKRLLNTDNKWVEKAVGILAGTTFGTTVDLGINKLAGRDVDLLQVIKQN
LIESGLGQFFGEPIDVVTGAFLITATDFTLPDIREDLRVQRKYNSTSREAGLLGPGWSFSYE
CRLYCSGNRLHAKLDSGITAVFAWDGSHAVNVTRGCEWLELTGEDDGWRIYDGRNYKCYHYD
GQGLLTAAEDRNGQCVRLYYEGERLTRITTPLGYSLDVEIRDGRLVQIRDHMGRTMQYRYEN
GFLSDVIHMDEGVTHYEYDSNGYLERAVDQAKVTYLENRYDDAGRVVLQTLANGDTYRADYH
PEKNRVTIVSSVHDKAVEHWYNEFGEILETSYQDGTKERYEYGENGHRTSRTDRLGRKTTWT
YDEAGRLTEEVQPDGLRTVHRYDAAGNEILRTDSAGRETAFEYDGHHNRTAERRTDGLQVRE
NRSVYDWMGRLTETADAEGNRTQYQYGEAGGKPSVIRFADGETCSFEYDKAGRMMAQEDACG
RTEYGYNARNKRALVRDGEGNETRWMYDGMGRLLALYLPKAWKEQHGEYSYSYDFLDRLIHT
KNPDGGHERLMRDGEGNVLKRVHPNAYDSCRDDGEGTTYDYDSDGNNIRIHYPDGGCERIFY
DSEGNRIRHVMPESYDPQTDDGEGFTYTYDACSRLTGVTGPDGVRQASYTYDPAGNLTEETD
AEGRCTYRSYTAFGELKEQLKPALEKDGVMLYERITWQYDRCGNVLLEQRHGGYWDSNGVLV
KEDGAGLALRFTYDSRNRRIRVEDGLGAVISCHYDVQGKLVYEEKAVSGEVRQVIHYGYDRA
GRLTERKEELDSGLAPLEGEPRYAVTRYRYDGNGNRTGIVTPEGYRILRSYDACDRLVSERV
VDDKNGIDRTTSVTYDYAGNITRIVRSGKGLGEWEQGYGYDLKDRIVHVKDCLGPVFSYEYD
KNDRRIAETLPQTGMTENGKSGYPKNQNRYRYDVYGRLLTRTDGSGTVQEENRYLPDGRLAL
SREADGQEIRYAYGAHGREEETGTARSRKAGRAAQKYRYDSRGRITGVVNGNGNETGYDLDA
WGRIQNIRQADGGEEGYTYDFAGNVTGTRDANGGVITYRYNSQGKVCEITDQEGNSETFRYD
REGRMVLHVDRNGSEVRTTYNVDGNPVLETGTDRNGENRVTRSFEYDASGNVRKAVAGGFCY
TYEYRPDGKLLKKSASGRTILSCSYHADGSLESLTDASGKPVFYEYDWRGNLSGVRDENGDM
LAAYAHTPGGKLKEICHGNGLCTRYEYDTDGNMIHLHFQRENGETISDLWYEYDLNGNRTLK
TGKCILSGDSLTDLAVSYRYDSMDRLTSESRDGEETAYSYDFCGNRLKKLDKSGTEEYHYNR
KNQLICRFSEKEKTAYRYDLQGNLLEAAGAEGTEVFSYNAFQQQTAVTMPDGKHLENRYDAE
YLRAGTVENGTVTSFSYHNGELLAESSPEGDTISRYIPGYGVAAGWNREKSGYHYYHLDEQN
STAYITGGSCEIENRYEYDAFGVLKNSMEEFHNRILYTGQQYDQTSGQYYLRARFYNPVIGR
FVQEDEYRGDGLNLYAYCKNNPVVYYDPSGYDSQYPCKEEMSAGAGESGRKTISLPEYDGTT
THGVLVLDDGTQIGFTSGNGDPRYTNYRNNGHVEQKSALYMRENNISNATVYHNNTNGTCGY
CNTMTATFLPEGATLTVVPPENAVANNSRAIDYVKTYTGTSNDPKISPRYKGN (SEQ ID
NO: 91)

NCBI Accession No. AVT32940.1 (TDD7)

MGDRLPAFVDGGDTLGIFSRGGIERDLASGVAGPASSLPKGTPGFNGLVKSHVEGHAAALMR
QNGIPNAELYINRVPCGSGNGCAAMLPHMLPEGATLRVYGPNGYDRTFTGLPD (SEQ ID
NO: 117)

NCBI Accession No. WP_189594293.1 (TDD8)

MSSRPFRKRLPGAVVRRWLGRGAVVASLSLLPQVVVPSGYDFAAQAQSVAARKKLEDRPEAK
INKVGVLRPGTSKAPKDKSAPASRKTRERLQEASWPKSGKATAAVTATSEATVNVGGLGMEL
TQEPAAPAAKSAKSTTKRKATGPAEKVTLRVHSRATAKKAGVNGVLLTVDPARGESNEKAED
TDKLRISLDYSSFSDVYGGNFGPRLSLVKLPACALTTPEKKSCRTQTPVAGADNEAESQTLT
GTVPARNLKAGTPMLLAAAADSSGGGGDFSATPLSPTATWEAGGSTGDFTWDYPLRVPPATA
GPSPNLSISYNSASVDGRTAGENNQTSLIGEGFSITESYIERKYASCKDDGQSGKGDLCWKY
ANATLVLNGKAVELVNACADKSACDTAALSEASGGTWKVKNEDGTRVEHLTGASGNGDNNGE
YWKVTDASGIQYYFGKHRMPGWSDKGTTDTADDDPSTYSTWAVPVFGDDSGEPCYKSSGFAD
SSCNQAWRWNLDYVVDTHDNASTYWYSKETNYYSKNADTTVNGTAYTRGGYLNRIDYGLRSD
LIYSKPAAQQVRFTYGQRCIVTNGCSSLTKDTKANWPDVPYDMICAANTKCTTQIGPSFFTR
QRLIDITTSVWTGTGTTRRDVDTWHLSHDFPDTGDASSPSLWLKSIQNTGKANTTTAAMPPI
VFGGIQMPNHVEGSGQDNLRYIKWRVRTIKSETGSTLTVNYSDPDCIWGSSMPSAVDKNTRR
CFPVKWSQSGTTPVTDWFHKYVVTSVLQDDPYGHSDTGETYYDYQGGAGWAYSDDEGLTKPS

NRTWSQWRGYGKVVTTSGNSEGPRSKKSTLYMRGLNGEKELDGTARVAKVTDSTGTAIDDSR
QYAGFVRETIAYNGSDELSGTINTPWSHKTGSHTYSWGTTEAWIVQAGETESRTKISTGTRT
VKQKTTYDTTYGMPITVEDSGDATKFGDESCVRTSYARNTSAWLVNRVSRTETYSVPCATIP
AIPADVVSDITTAYDAKAVGAAPTQGDITATYRVASYNAADKTPVYQQVSSSTFDKLGRPLT
ETNALDRTVKTSYVPDDTGYGPLTSKTTTDPKLYTSTTEVDPAWGAASKTTDANGNVTEWSF
DALGRLRSVWKPDRSRTLDDAASIVYAYSVNNDKETWVRTDALKADGKTYNSSYEIFDSLLR
PRQKQVPAPNGGRVISEMLYDDRGLAYIANSQVHDNSAPSGTLANTYTGSVPASTETVFDAA
GRATDSIFRVYGQEKWRTKTDQQGDRTAVTAAAGGTGTLTIVDARGRVTERREFGGPAPTGT
DYTRTLYEYTPGGQIKKMTGPDGAVWTYEYDLRGRKTTSTDPDKGSITTTYNDADQPLTATT
TLDNVSRTLINDYDELGRPTGTWDGTKDNAHQLTKFTYDSLAKGQPTASIRYVGGTTGKIYS
QSVTGYDALNRPKGTKTVIAATDPLVTAGAPQTFTTSTAYNIDGTVQSTSLPAAAGLPAETV
KNTYNSLGMLTGTDGMTDYVQHIGYSPYGEIEETRLGTSTEAKQLQVLNRYEDGTRRLANTH
TLDQTNAGYTSDVDYVYDATGNVKSVTDKANGKDTQCFAYDGYRRLTEAWTPSSNDCATARS
ASALGGPAPYWTSWTYKPGGLRDSQTEHKTSGDTKTVYGYPAVNTSGTGQPHTLTSVTVGSG
SAKTYTYDEHGNTTKRYSPTGTAQSLTWNIEGELTRLTEGTKTTDYLYDANGELLIRRSPDK
TVLYLGGQELHYDTATEKFTAQRYYPAGDATAVRTETGLSWMVDDHHGTASMTVDATTQAVT
RRYTKPFGEARGTAPSVWPDDKGFLGKPADTGTGLTHIGAREYDPTLGRFLSVDPVLAPDDH
ESLNGYAYANNTPVTLSDPTGLRPDGMCGGSSSSCGGGTETWTLNSKGGWDWSYTKTYTKKF
TYRTGNGGTRTGTMTTTVRTEVGHKAVRIVFKKGPEPKPAKKDGQCSSCWAMGTNPGYSPGA
TDDWIDRPKLETWQKVVLGAISVVAAGVILAPAAIVVGEGCLAAAPVCAAEIAEAATGGASG
GSAVVGAGVVATGAKAVTTGKSLSESQATLSVAQRLLATIGEEGKTAGVLELDGELIPLVSG
KSSLPNYAASGHVEGQAALIMRDRGATSGRLLIDNPSGICGYCKSQVATLLPENATLQVGTP
LGTVTPSSRWSASRTFTGNDRDPKPWPR (SEQ ID NO: 118)

NCBI Accession No. TCP42004.1 (TDD9)

MAFGIGTSRRGSGGGRGWGRRLVTPVAALALLAPLGEAQDAVAQDAGAVRSGPVQPDVPKPR
VSKVKEVKGLGAKKARDRVAAGKKAGAAQAARARREQTAVWPGPDTASIELADDRRAKAELG
GASVSVVPENGRKTAASGTAQVTILDQKAADKAGVTGVLLSATADTAGTAEVSVDYSGFASA
FGGDWAQRLHLVQLPACVLTTPEKAVCRRQTPLKTDNNASEQSVAAQVALAKAEPGAPSAQS
VASAEGPSATVLAVTAAAAGSGASPKGTGDYAATELSPSSAWEAGGSSGAFTWNYGFTVPPA
AAGPTPPLALSYDSGSIDGRTATTNNQGSAVGEGFSLTESYIERSYGSCDKDGHADVWDHCW
KYDNASIVLNGKSNRLIKDDTSGKWRLETDDSTVTRSTGADNGDDNGEYWTVTTGDGTKYVF
GENKLDGAADQRTNSTWTVPVFGDDSGEPGYDKGDTFAERAVTQAWRWNLDYVEDTSGNAST
YWYAKDSNYYPKNKATTANASYTRGGYLKEIRYGLRKDALFTDDADAKVVFAHAERCTVGSC
TTLTKDTAKNWPDVPFDAICSSGDSECNAAGPSFFSRKRLTGISTFSWNAASKAFDPVDTWE
LTQDYYDAGDIGDTTDHVLVLESIKRTAKAGATAIDVNPVTFTYQLRPNRVDGTDDILPLKR
HRIETITSETGSITTVTLSQPECKRSTVLDAPQDSNTRPCYPQFWNINGATKASVDWFHKYR
VLAVAVDDPTGHNESIEHAYDYAGAAWHYSDDPFTTKNERTWSEWRGYRDVTTYTGALDTTR
SKSVSRYMLGMDGDKNTDGTTKSVSTAPLMDTDVDFAALTDSDPYSGQLLQQVTYSGSQPIS
TSYTNFTHKNTASQTVPDATDHTARWVRPNSSYASTYLTASKTWRTQVTTSRYDDLGMVTSH
DDYGQKGLSGDEICTRTWYARNTEAGINSLVSRTRTVGKECSVDDTALDLPADNKRSGDVLS
DTATAYDGATWSDSMKPTKGLVTWTGRAKGYASGTPSWQTLTSAAPADFDVLGRPLKVTNAE
GQPTTTAYTPVTAGPVTKIISGNPKGFKTTSFLDPRTGQELRTYDANLKKTERVYDALGRLT
QVWLPNRDRGSESATFGPSVKFEYTIDNNDPSWVSTAALKKDGKTYATSYAIYDAMLRPLQS
QTETSNGGRLLTDTRYDTRGLPYETYANIFDTTSTPNGTYTRAEYGEAPNQNATVFDGAGRP
TKSTLLVFGVEKWSTTTSYTGDSTATTALDGGTASRAITNIRGHTVESREYAGKSPADAQYG
DGLGVGFASTRTLYTRGGLQKQITGPDDATWSYTYDLFGRQVEAEDPDKGTSSTEYDVLDRA
TKSTDSRSKSILTAYDELGRMVGTWAGSKTDTNQRTEYTYDKLLKGQPDKSIRYVGGKAGQA
YTDTITEYDSLSRPVAASLELPADDPFAKVGALGSASRTLSFRHAYNLDDTVKTAEEPALGG
LPSEIIDYGYNNVGQVTSVGGSTGYLLGATYSPLGQPWEQLLGTANTADHKKVSIRNTFEDG
TGRLTRSNVKADSQPYMLQDLNYSFDQVGNVTSITDPTTLGGTSSAETQCFTYDSHRRLTEA
WTPSQQKCSDPRSTSSLSGPAPYWTSYTYNTAGQRTTETTRKAAGDTTTTYCYTKTDQPHFL
TGTTTKGDCATRERTYTPDTTGNTTKRPGASTTQDLAWSEEGKLTKLTENGKATDYLYDATG
ELLIRNTTSGERVLYTGTTELHLRTDGTTWAQRYYAAGDQTVAMRSNESGTNKLTYLAGDHH
GTSSLAISADSTQTVSKRYMTPFGAERGKPTGTAWPDDKGFLSKTTDKTTGLTHIGAREYDP

AIGQFISTDPILDPAQPQSLNGYSYANNTPVTAADPSGLWCDSCNDGKGWTRPDGGTRGDEN
GGKNPDGSVRGTPGFPSTRPTTVGYGNSPGAGKVITDLGSGTPALPPPDVYQDYQPKLPGVG
QMGRNGTYMPELSYELNVELYFRERCSFSWTEECESIRAFYTHGEDSHGLPRYWTDVQDIPT
VNTCPICENIGFDIILATLPIGKVGKLRFAPKVESAESMLRSLSQEGKTAGVLDINGELIPL
VSGTSSLKNYAASGHVEGQAALIMRERGVASARLIIDNPSGICGYCRSQVPTLLPAGATLEV
TTPRGTVPPTARWSNGKTFVGNENDPKPWPR (SEQ ID NO: 119)

NCBI Accession No. WP_171906854.1 (TDD10)

MRGWVRAVSIPVIVGVLSTALSMPPSFADQEPVARTEATTDGLPTNADEGQRAEPPALIPSE
NRIPGVGLKSEIESQPTAASVADGPLPSERSDSFFPALAPTPPTIVGYVPTSLAPGCAEWGA
LRWTHPDSRPNGLVHLYTFELYRDSDDAMVWDQLFDYTLTGAGVVSDVAGDCESILPDPQAT
PIVELGESYYAKVYAWDGTGWSAPATSSAYPAVALPGLTDEAARGVCVCDTSTGRLYPLNIL
RADPVNTATGTLTESATDLTIPGVGPAISASRTYNSTDPTVGPLGKGWSFPYFSELESAASS
VTYKAEDGQEVEYALQGGAYRLPPGASTRLRSVSGGYQLETKSHQVIGFDQNGRLEYARDSS
GQGVSLAYATNGTLDKITDASGREVDVTMDASGKVTAIALSDGRSVSYGYTGDLLTSVTDVR
GGVTEYEYDAAGRLAAITDPLGNEVMRSTYDAQGRVISQVDAGGGTWGFEYVDDGAYQTTRT
TDPRGGVSRDVYYNNVLVESETAGGAITTYQYDERLRLAATVDPHGRTTRHTYDANDNLLST
THPNGDREAFTYSSGGDLLTETSPEGRKTTYTYDANHRVATTTDPNGGVTSYTYNTDGQVLT
ETSPEGNVTEFEYDAQGNRVATISPEGRRTTATFDAYGRLESQTTARGHVAGADPADFTTTF
AYDVASNLTSSTDPLGHVTEYEYDLNNRRTTVIDPLDRRTETEFDAAGRVVKIIEPGGAETV
HEYDLAGNQVATTDAEGGRTTRTFDLDAHMITMTAARGNEPGAEPADFTWGYEYDGLGNVVE
ETDSAGGIVSYGYDERYRQTSVTNQANETTTTAYDGDGNTVSVTDPLDRTVSTTYNGLNLPA
TVTDPAGKVSTVIYDRDGNRTSTTTPLGHKATFTYDGDGMLVQDQTPNGNGRISTYTYDADG
NQIRTVDPQGRFTTATFDNAGRVSSRSLWNVTTTYGYDDAGRLTTVTGGDGAVTEYGYNTAG
DLVTVTDPNDHVTTHTYDDAHRRTATTDALNRTRTFGYDADGNQTSTVLARGPASGDLARWT
VTQSYDELGRRTGVTTGSTASTASYAYDPVGRLTGVTDAGGTTTTVYDDAGQIASVTRGSQA
YGYTYDPRGMVKTITQPGGVTVTNTFDDDGRLATTASTNAGTTAFSYDKNNNLTRIDNLAAT
GLVNRWQQRNYDRADALVSTTTGTGTTTDPTQTVTYSRDGAGRPFVIRRGAGGTQAPGEAHF
FDAAGRLAQVCYDASSMFGQNCATADETLAYTYDGAGNRLTETRTGGTTPGTTTYTYDAANQ
LTQRGNTTYSYDADGNQISDGATSWTYDELNRLVGIDTPTADSQLTYDGLGNRTSVTTGATT
RTFSWDINNPLPLLTSVTQGTSTTRYRYGPDAIPVNANINGTNHALLTEDLNSLTTTYNRTT
GAKSWTTTYEPFGTPRNTTSTGLTTAQVGLGYTGEYLDPTTGLLNLRARNYNPTLGQFTSTD
PVETPQGTPSISPYAYVDNRPTVLTDPSGACFFIDMPWIPGCSEPSWADEVTPATNGVLAGL
ISAAEDTFYLTGMALGVDWVGYDGDLAQQLFDEAAVEGNYHGETYQQAQLVGGLVALVGGAA
STAASLARICTSLVRKIRPPVASGGLATEVPAYAGSRTAGTLVTPDGAEFPLISGWHPPAAS
MPQGTPGMNIVTKSHVEAHAAAIMRNQGLSEATLWINRAPCGGKPGCAAMLPRMVPSGSTLT
INVVPNGSAGSIADTLIIRGIG (SEQ ID NO: 120)

NCBI Accession No. WP_174422267.1 (TDD11)

MSDSENRLTRASDSPASGKTQSESKVNTACDSLLDTAGSTYDSLKQPFSSKGGALHHVSEAV
NALASLQGAPSQLLNTGIAQIPLLDKMPGMPASVISAAHLGTPHAHSHPPSDGFPLPSMGAT
IGSGCLSVLIGGLPAARVQDIGIAPTCGGLTPYFNIETGSSNTFIGGMRAARMGIDMTRHCN
PMGHAGKSGEEAEGAAEKGEQAASEAAEVSSRARWMGRAGKAWKVGNAAVGPASGVAGAASD
AKHHEALAAAMMAAQTAADAAMMLLSNLMGKDPGIEPSMGMLMDGNPTVLIGGFPMPDSQMM
WHGAKHGLGKKVKARRADRQKEAAPCRDGHPVDVVRGTAENEFVDYETRIAPGFKWERYYCS
GWSEQDGELGFGFRHCFQHELRLLRTRAIYVDALNREYPILRNAAGRYEGVFAGYELEQRDG
RRFVLRHGRLGDMTFERASEADRTARLVNHVRDGVESTLEYARNGALMRIDQEKGPGRRRQL
IDFRYDDCGHIVELYLTDPQGETKRIVHYRYDTAGCLAASTNPLGAVMSHGYDGRRRMVRET
DANGYSFSYRYDSQDRCIESMGQDGLWHVSLDYQPGRTVVTRADGGKWTFLYDEARTVTRIV
DPYGGTTERVSGDNGRILREIDSGGRVMRWLYDERGGNTGRMDRWGNRWPTKDEAPVLPNPL
AHTVPNTPLALQWGDARHEDLADTLLLPPEIAKIAASFFPPQPFSASTEQCDETGRVIARTD
GYGQAERLRLDATGNLLQLCDRDGRDYCYSIASWNLRESETDPLGNTVRYRYSPKQEITAIV
DANGNESTYTYDYKSRLTSVTRHGTVRETYAYDVGDRLIEKRDGTGNALLRFEVGEDGLQKT
RILASGETHTYKYDHRGNFTRASTDKLDVTLTYDAYGRRTGDKRDGRGIDHSFVGGRLESTT
YFGRFVVRYEAGQAGDVMIHTPGGGIHRLRRAADGTVLLRLGNRTNVLYGFDADGRCTGRLS

WPEGRTAEIHCVQYRYSAVGELRCVIDSTGGTIEYQYDAAHRLVGESRDGWAVRRFEYDQGG
NLLSTPTCQWMRYTEGNRLSSASCGAFRYNSRNHLAEQIEENNRRTTYHYNSMDLLVQVKWS
DRQESWRSEYDGLCRRIAKAMGQARTQYFWDGDRLAAEAAPDGRLRIYVYVNEASYLPFMFI
DYPSCDAEPESGSAYYVFCNQVGLPERIESAMGLDAWRAEEIEPYGSIRVATGNAIDYDLRW
PGHWFDVETGLHYNRFRYFQPTLGRYLQSDPAGQSGGVNLYAYSANPLVFVDVLGLECPHND
KSTTECARCEAKEEVDQREKRDKELAREIYHIEDKYSDSHAGIGLDPDEKKRALEDKIDYDD
LVRKREKAREDLLEAEKRLREEEIRAKYPTPEEAQLPPYDGDTTYALMYYTDEHGKSHVVEL
SSGGADDEHSNYAAAGHTEGQAAVIMRQRKITSAVVVHNNTDGTCPFCVAHLPTLLPSGAEL
RVVPPRSAKAKKPGWIDVSKTFEGNARKPLDNKNKKST (SEQ ID NO: 121)

NCBI Accession No. WP_059728184.1 (TDD12)

MSEPANRLTRASEPSERHAAQSESKADTACESLLGTVKSTFDPFKQTFSSDGSALHHVSEAV
NALASLQSAPSQLLNTGIAQIPLLDKMPGMPAATIGVPHLGTPHAHSHPPSSGFPLPSIGAT
IGSGCLSVLIGGIPAARVLDIGIAPTCGGLTPYFDIQTGSSNTFFGGMRAARMGIDMTRHCN
PMGHVGKSGGKAAGAAEKTEEAASEAAQVTSRAKWMGRAGKAWKVGNAAVGPASGAAGAAAD
AAHGEELAAAMMAAQTAADAAMMLLGNLMGKDPGIEPSMGTLLAGNPTVLVGGFPLPDSQMM
WHGVKHGIGKKVRARIANRRKEVSPCTDGHPVDVVRGTAENEFVDYETKIAPAFKWERYYCS
GWSEQDGALGFGFRHCFQHELRLLRTRAIYVDALNREYPILRNAAGRYEGVFAGYELEQRDG
RRFLLRHGRHGDMTFERENEADRTARFVSHVRDDVECTLEYARNGALARIAQEDARGLRRQL
IDFRYDDRGHIVELCLTDPRGQTRRLAHYRYDAAGCLTVVTDPLGAVTSHGYDDRRRMVRET
DANGYSFSYRYDSQGRCIETVGQDGLLHVVLDYQPGRTVVTRADGGKWTFLYDNARTVTRIV
DPYGGMTERVIGGDGRILREIDSGGRVMRWLYDERGRNTGRMDRWGNCWPTRDEAPVLPNPL
AHTVPVTPLDLQWGEVSPAELTDSVLLSPEIQKVAESLFQQPAFSPSEQHDARGQVVARTDE
HGGVERFRRDAAGNIIQVCDKDGRAHHYGIASWSLRESETDSLGNTVRYRYSNKQEITSIVD
ANGNESAYTYDYKGRITSVMRHGVVRETYTYDAGDRLIEKRDGAGNLLLRFEVGENGLHSKR
ILASGETHTYEYDRRGNFTKASTDKFDVTRTYDAHGRRTGDKRDGRGIEHVYGDGRLCSTTY
FERFTVRYEAEADGEVLIHAPVGGTHRLQRSSDGQILLRLGNGANVLCRFDAHGRCVGRLVW
PEGRPKECHRVAYQYSAMGELRRVIANTTGTTEYLYDDAHRLIGESHDGWPVRRFEYDCGGN
LLSAPTCQWMRYTEGNRLATASRGAFYYNDRNHLAEQIGENNHRTSYHYNSMDLLVKVTWSD
WPEVWTAEYDGLCRRIAKAMGPARTEYYWDGDRLAAEIAPNGQLRIYVYVNETSYLPFMFID
YDGCDAAPESGRGYYVFSNQVGLPEWIEDIAGACVWRAMEIDPYGAIRVAPGNELGYNLRWP
GHWLDPETGLHYNRFRSYHSALGRYLQSDPAGQSGGINLYAYTANPLVFVDVLGRECPHLNE
SSSECSQCENREEAERIRKEMLQSISRRMDIEGDVTGHPGILLTQAELTGKYSHYAEEYKQL
LKDIDTKREAEEAALLREAYPSMEGATLPPFDGKTTIGLMFYTDASGQYQVKKLFSGEKVLS
NYDATGHVEGKAALIMRNEKITEAVVMHNHPSGTCNYCDKQVETLLPKNATLRVIPPENAKA
PTSYWNDQPTTYRGDGKDPKAPSKK (SEQ ID NO: 122)

NCBI Accession No. WP_133186147.1 (TDD13)

MSTPPGNPASPANEPPPPPAPLISPTGNTSVDALASAVNAGAQPFQQLGNPKANTLDRVTNV
VSGAVGSLGALDQLLNTGMAMIPGANLVPGMPAAFIGVPHLGVPHAHAHPPSDGVPMPSCGV
TIGSGCLSVLYGGMPAARVLDIGLAPTCGGLAPIFEICTGSSNTFIGGARAARMALDLTRHC
NPLGMSGAGHAEQDAEKASALKRAMHIAGMAAPVASGGLTAADQAVDGAGAAAVEMTAAQTA
ADAIAMAMSNLMGKDPGVEPGVGTLIDGDASVLIGGFPMPDALAMLMLGWGLRKKAHAPEGA
GEPKRTEQGECKGGHPVDVVRGTAENQFTDYATLDAPEFKWERYYRSDWSERDGALGFGFRH
SFQHELRLLRTRAIYVDGHGRAYAFGRSASGRYEDVFAGYELEQQGENRFVLLQATRGEFTF
ERASAAQASARLVRHVHEGVESALRYAGDGTLRHIEQTAQREQRHRMIDLLYDARGHVVEMR
VTDPRGAVLCAARYRYDATGCLVASTDALGASMTYGYDAWRRMIRETDANGYAFSYRYDSDG
RCVESAGQDGLWRVLLDYQPGRTVVTQADGGRWTYLYDAARTVTRIVDPYGGATERVIGDDG
RIVEEVDSGGRVMRWLYDERGENTGRQDRWGNRWPTRDEAPVLPNPLAHVVPARPLELLWGD
ARPEDFTDRLLLPPEIEAVAAAAFAPSAAVPKPAEQRDGAGRVIRRTDESGHAECLHRDAAG
NVVQLRDKDGRYYGYAIASWNLRESETDPLGNTVRYRYSSKQNITAVVDANGNESRYTYDYK
SRLTRVARHDTIRESYVYDTGDRLIEKRDGAGNTLLRFEVGENGLHSKRILASGETHTYEYD
RRGNFTRASTDKFEVTLTYDAFGRRTGDKRDGRGVEHSFVGQRLESTTWFGRFVVRYETGPS
GDVMIHTPGGDVHRLQRAADGTVLLRLSNSTNVLYKFDENGRCAGRLTWPDGHTSANRCVQY
RYSAVGELRQVIDSKGGTTEYQYDDAHRLVGESREGWAFRRFEYDRGGNLLSTPTCQWMRYT

EGNRLSGAACGAFCYNSRNHLAEQIGENNRRTTWHYNSMDLLVRVQWSDRQENWSAEYDGLC
RRIAKAMGQARTQYFWDGDRLAAEVAPNGQLRIYAYVNETSYLPFMFIDYDGCDAAPESGRT
YYVFCNQVGLPEWIEDISRGCVWGVNEIDPYGAICVAPDNELEYNLRWPGHWFDPETDLHYN
RFRSYSPVLGRYLQSDPAGQAGGINLYAHTANPLVFIDVLGRECPHGNESSSECSQCADREE
AERINAKILQLISKKMSIEDAVTGHPGELIPLPHFEIDKEYSHYAKEYKQLLADIDALAEAR
EDALLREQFPSMDAVTLPPFDGKTTIGYMFYTDANGQYHVRKLYSGGKVLSNYDSSGHVEGM
AALIMRKGRITEAVVMHNHPSGTCHYCNGQVETLLPKNAKLKVIPPANAKAPTKYWYDQPVD
YLGNSNDPKPPS (SEQ ID NO: 123)

NCBI Accession No. WP_083941146.1 (TDD14)

GSSGKNVRMPRDYASELPEYDGKTTHGVLVTNEGKVIQLRSGGKEEPYTGYKAVSASHVEGK
AAIWIRENGSSGGTVYHNNTTGTCGYCNSQVKALLPEGVELKIVPPTNAVAKNAQARAVPTI
NVGNGTQPGRKQK (SEQ ID NO: 124)

NCBI Accession No. WP_082507154.1 (TDD15)

MDAETGLVYFQARYYDPQLGRFITQDPYEGDWKTPLSLHHYLYAYANPTTYVDLNGYYARDA
NEVQRYIIAESNCAKTGSCDAVTALREPSEARQRSAANCKSLDRCREIADDAARSEGDISAR
IKALQKDLRNGIEANPTTGIKTIWELDKQLEARNISAGAVREAGRHVRWRAFVENRELTDHE
KVAPAAEMYGVLSGGRIVIARAVARSSVTRASITQESKTIGVTAEVAPNESLRNTSGDLRAS
ANSARNQPYGNGQSASASPSTNSAGSSGKNVRLPRDYASELPEYDGKTTYGVLVTNEGKVIQ
LRSGGKEVPYSGYKAVSASHVEGKAAIWIRENASSGGTVYHNNTTGTCGYCNSQVKALLPEG
VELKIVPPANAVARNSQAKAIPTINVGNATQPGRKP (SEQ ID NO: 125)

NCBI Accession No. WP_044236021.1 (TDD16)

MLASTWLDLVIGVDLHFELVPPVMAPVPFPHPFVGLVFDPWGLLGGLVISNVMSVATGGSLQ
GPVLINLMPATTTGTDAKNWMLLPHFIIPPGVMWAPMVRVPKPSIIPGKPIGLELPIPPPGD
AVVITGSKTVHAMGANLCRLGDIALSCSDPIRLPTAAILTIPKGMPVLVGGPPALDLMAAAF
ALIKCKWVANRLHKLVNRIKNARLRNLLNRVVCFFTGHPVDVATGRVMTQATDFELPGPLPL
QFERVYASSWADRASPVGRGWSHSLDQAVWLEPGKVVYRAEDGREIELDTFELPGRMLQPGQ
ESFEPLNRLLFRCLDGHRWEVESAEGLVHEFAPVAGDADPAMARLTRKRSRQGHAITLHYDG
KGCLTWVQDSGGRIVRFEHDEAGHLTQVSLPHPTQPGWLPHTRYIYSPEGDLVEVVDPLGHR
TRYEYVGHLLVRETDRTGLSFYFGYDGTGPGAYCIRTWGDGGIYDHEIDYDKVNRVTFVTDS
LGATTTYEMNVANAVVKVIDPRGGETRYEYNDVLWKTEEVEPAGGATRYEYDARGNCTKSTG
PDGATVQVEYDARNVPIRAVNPCGEEWQWVYDAQGQLVERIDPLGETTRYEYDKGMVVTITE
ASGVTTAEYDDSRNLRRVQGPSEAETSYVYDALGRMVVKRSPARVAERLHYDACGRLVTVEQ
PDGNVWRLAYDGEGNLTEIQDHHQRVRMRYGGYHQMVSRQEAEDTTLFRYDSEGRLVAIENE
AGEIYQYELDSCGRAGLERGFDGGCWKYERDAAGRVIKLRKPSGAEARLIYDAMGRLVEVRR
SDSAVERFRYRKDGALIEAENSTIQVKFERDALGRVVREMQGGHWVESSYERGARTWVASSL
GVHSAIMRDERRSVVAMTAGRGVDEWRVELSRDAFGLETERKLSSGIVSTWARDALGRPRHR
GVAHSNNVLFGVEYQWAPGSRLVALIDTERGTTAFHFDERSRLVGAKLPGGRIDRREPDRIG
NIYRAQDQRDRTYSDGGILRGAGETRYTHDLDGNLTQKVLPDGATWSYSYNAAGCLKEVERP
DGTRVTFAYDALGRRVSKRWGENEVWWLWDRHVPLHEISTRAEPITWLFEPESFAPIAKIEG
DRHYDILCDHLGAPTVVLDEAGVVTWRARLDIHAAVQPEIAETECPWRWPGQYEDQETGLYY
NRFRYYDPEADRYISQDPLGPVGGLNLYSYAADPLTWSDPLGLQPDPPPPPTPMGNTLPGWD
GGKTQGWFVYPDGTERHLISGYDGPSKFTQGIPGMNGNIKSHVEAHAAALMRQYELSKATLY
INRVPCPGVRGCDALLARMLPEGVQLEIIGPNGFKKTYTGLPDPKLKPKGCS (SEQ ID
NO: 126)

NCBI Accession No. WP_165374601.1 (TDD17)

MTACSDSPRLPPSLLELPDTPCPEPDEAASPFPAELPHSATVEAGAIAGSFGVTSTGEATYT
IPLVVPPGRAGMQPELAVQYDSASGEGVLGMGFSVTGLSAVTRCPRNLAQDGEIRAVRYDEG
DALCLDGKRLVEVGGGGEVVEYRTVPDTFARVVASYEGGWDRARGPKRLRVFTRAGRVLEYG
GEPSGQVLAKGGVIRAWWATRVSDRSGNTIDFHYQNETSASEGYTVEHAPRRIEYTGHPRAA
ATRAIEFVYAPRRPGTGRVLYSRGMALRSSQQLDRIRMLGPGGALVREYRFSYTSGPATGRR

LLNAVRECAADGRCKPATRFRWHHGTGPGFAEVGTRLRVPESERGSLMTMDATGDGRDDLVT
TDLDLPVDDDNPITNFFVAPNRMAEGGSSSFGALALAHQEMHHAPPSPVQPELGTPIDYNDD
GRMDIFLHDVHGRYPDWHVLLATPEGTFRRKSTGIRRKFGIDAPPPLDLNSRNASAHLADVD
GDGIADLLQCEDTGSVFTDWTLHLWRPAASGFEPEPSRIPALRGHPCNAETHLADVDSDGKV
DLLVYEATITGNGTLFGTTFEALSFVRPGEWTKRATGLPVLKAGSGGRVIVLDVNGDGLPDA
VETGFDDGQLRTFINTGDGFAAGVSSLPSFVFDADAFAKLAAPIDHNSDGRQDLLMPIREPG
GPVLWKILQATGSTGDGTFAVIDARLPVSEVLVDREITLAHPWAPRVTDVDGDGNQDVVLAV
GKELRVFRSRLREEDLLWTVSDGMSAYDPEEAGHVPKVQIEYSHLSAAEPGVRGEQRTYLPR
YDTGEPGDGACDYPVRCALGPRRVVSRYAVNNGADRLRTFQVAYRNGKYHRLGRGFLGFGVR
IVRDAASGAGSAEFFDNVTFDPSDRSFPLAGHVVREWRWTPEPQQKGVSRVELSYTERLIHA
ILTNRGKSYFTLPVYQKQRREQGEHRRDSGKTLEEYVRDTWYAPTQVVSRTERLVSAWDAFG
NIREESTSTAGVDLTLKVKRTFRNDEDAWLIGLLETQQECSRALSIEQCRTSSRAYDRHGRV
RTESAGSDDDDPETVVRVRYTRDAFGNVIHTRAEDAFGGRRKACVSYDAEGVFPYAQRNPEG
HVTYTRYDAGHGALEAVVDPNGLATQWAHDGLGRITEERRPDGTTTRATLSRTRDGGPRGDA
WRVLRRTATDGGADETVELDGFGRPIRGWAYKARTDDGPAERVVQEIAFDQSGERVARRSLP
AAEGTPRERMQVETYGHDATGRIAWHRAAWGAETRYRYLGRTVEVEGPGGRVTTIENDALGR
PVRIVDPEGGVTSYAYGPFGGLWTVTDPGDAKTTTERDAYGRVRRHIDPDRGTAVAHYDGFG
QQTSTVDALGREVSWKHDRLGRAVERSDEDGTTTWTWDEAEHGVGKLAEVASPEGHRTTYRY
DALGRLREEELAIEGERFATTVDYDGHSRPFRLWYPQAEGERRFGVRRIFDAHGHLVGLRNE
RSREMFWRLEDTDEAGRIRIEEFGNGVTTERSYHETKGRLRRVATMKDHVVLQDLWYGYDDR
LNLSSRRDDRLERTEHFRYDKLDRLTCAARHERFCLFETTYAPNGNIREKPDVGEYTYDPEH
PHAVRTAGADVFAYDAVGNQVRRPGVEEIRYTAFDLPASITLAGGTGTVDLDYDGDQRRIRK
TTPMEQTVYAGDLYERVTDLATGVVEHRYTVRSSERAVAVVTKRAGGEARTLYIHVDHLGSV
DLLTEGRGEDAGREVERRSYDAFGARRDPVTWRRAPKAEAPPALLARGFTGHGSDDELGLVH
MKGRLYDPKIGRFTTPDPVVSRPLFGQSWNAYSYVLNNPLAYVDPSGFQEAVPEDRGGSSRA
AGAEFTSDELGLPPIEELVVARFPEHEARSDADANAMGAEVGGAVPPVDVGVYGTSAGFVPQ
PGPSSPEHASAASVVGEGLLGAGEGTGELALRVARSLVLSALTFGGYGTYELGRAMWDGYKE
NGVVGALNAVNPLYQIGRGAADTALAIDRDDYRAAGAAGVKTVIIGAATVFGAGRGLGALEE
ATTAAGIARGAPSLPVYTGGKTTGVLRTATGDMPLVSGYKGPSASMPRGTPGMNGRIKSHVE
AHAAAVMRERGIKDATLHINQVPCSSATGCGAMLPRMLPEGAQLRVLGPDGYDQVFIGLPD
(SEQ ID NO: 127)

NCBI Accession No. NLI59004.1 (TDD18)

MVIIGRIDTNESTVSLYQWSLLPATDTNCYKEITVEQYKNNQLVRKVSFSKAFVVNYTESYS
NHVGVGTFTLYVRQFCGKDIEVTSQELNSVSNLTPNLPNSVEKDVEVVEIAEKQAVVKSDTS
NLKQSNMSITDRLAKQKEKQDNTNIIDNRPKLPDYDGKTTHGILVTPNSEHIPFSSGNPNPN
YKNYIPASHVEGKSAIYMRENGITSGTIYYNNTDGTCPYCDKMLSTLLEEGSVLEVIPPINA
KAPKPSWVDKPKTYIGNNKVPKPNK (SEQ ID NO: 128)

NCBI Accession No. KAB8140648.1 (TDD19)

MLYAYGPESVVAERTIVGTTVADAGKAAFRVLDDTLAEGVEHSANKADEAGELIEAVVEQCL
RNSFSADTLVTTASGLRPISTIAVGELVLAWDATTRSTGYYPVTAVMLHTDAAQVHLSVGGE
HVETTPEHPFYTLERGWVAAGDLWDGAHVRRADGSYALTLVLWLDAEPQVMYNLTVATAHTF
FVGVERALVHNAGCPGDALPPYGTKGSKTTGILDTGNESILLESGENGPGMMVPRDTPGMSG
AMPNRAHVEGHTAAIMRNENIRLADLYINRMPCSGAYGCMVNLPHMLPEGSILRIHVRAKLS
DPWTTLPPFVGISDTLWPPSGLNPKIVLP (SEQ ID NO: 129)

In some embodiments, said sequences do not include a signal sequence, if present.

[0061] In some embodiments, the cytidine deaminase may comprise the toxic domain of a TDD. Examples of toxic domains for TDD1-TDD19 are as follows: TDD1 (SEQ ID NO: 92), TDD2 (SEQ ID NO: 95 or 134), TDD3 (SEQ ID NO: 98), TDD4 (SEQ ID NO: 101 or

143), TDD5 (SEQ ID NO: 104), TDD6 (SEQ ID NO: 107 or 152), TDD7 (SEQ ID NO: 157), TDD8 (SEQ ID NO: 162), TDD9 (SEQ ID NO: 167), TDD10 (SEQ ID NO: 172), TDD11 (SEQ ID NO: 177), TDD12 (SEQ ID NO: 184), TDD13 (SEQ ID NO: 189), TDD14 (SEQ ID NO: 194), TDD15 (SEQ ID NO: 199), TDD16 (SEQ ID NO: 204), TDD17 (SEQ ID NO: 209), TDD18 (SEQ ID NO: 214), and TDD19 (SEQ ID NO: 219), e.g., as shown in **Table 9**. The toxic domains of TDD1-TDD19 may be split into half domains, e.g., as shown in **Table 9**. In some embodiments, the toxic domains of TDD1-TDD19 are split into half domains at the residues indicated in **Table 9**. In certain embodiments, TDD half domain pairs may comprise the amino acid sequences of SEQ ID NOs: 93 and 94, SEQ ID NOs: 96 and 97, SEQ ID NOs: 99 and 100, SEQ ID NOs: 102 and 103, SEQ ID NOs: 105 and 106, SEQ ID NOs: 108 and 109, SEQ ID NOs: 130 and 131, SEQ ID NOs: 132 and 133, SEQ ID NOs: 135 and 136, SEQ ID NOs: 137 and 138, SEQ ID NOs: 139 and 140, SEQ ID NOs: 141 and 142, SEQ ID NOs: 144 and 145, SEQ ID NOs: 146 and 147, SEQ ID NOs: 148 and 149, SEQ ID NOs: 150 and 151, SEQ ID NOs: 153 and 154, SEQ ID NOs: 155 and 156, SEQ ID NOs: 158 and 159, SEQ ID NOs: 160 and 161, SEQ ID NOs: 163 and 164, SEQ ID NOs: 165 and 166, SEQ ID NOs: 168 and 169, SEQ ID NOs: 170 and 171, SEQ ID NOs: 173 and 174, SEQ ID NOs: 175 and 176, SEQ ID NOs: 178 and 179, SEQ ID NOs: 180 and 181, SEQ ID NOs: 182 and 183, SEQ ID NOs: 185 and 186, SEQ ID NOs: 187 and 188, SEQ ID NOs: 190 and 191, SEQ ID NOs: 192 and 193, SEQ ID NOs: 195 and 196, SEQ ID NOs: 197 and 198, SEQ ID NOs: 200 and 201, SEQ ID NOs: 202 and 203, SEQ ID NOs: 205 and 206, SEQ ID NOs: 207 and 208, SEQ ID NOs: 210 and 211, SEQ ID NOs: 212 and 213, SEQ ID NOs: 215 and 216, SEQ ID NOs: 217 and 218, SEQ ID NOs: 220 and 221, or SEQ ID NOs: 222 and 223.

[0062] As used herein, unless specified otherwise, the term "TDD" refers to the TDD toxic domain.

[0063] Where the present disclosure refers to a cytidine deaminase (e.g., a TDD described herein), it is contemplated that other cytidine deaminases can be used in the fusion proteins and cell editing systems described herein. The cytidine deaminase can comprise wild-type or evolved domains. In certain embodiments, the cytidine deaminase may be, e.g., apolipoprotein B mRNA-editing complex 1 (APOBEC1) domain or an Activation Induced Deaminase (AID).

[0064] The present disclosure also provides other potential cytidine deaminases. Such cytidine deaminases may be used, e.g., in the fusion proteins and cell editing systems described herein. In some embodiments, the cytidine deaminases are functional analogs of a

TDD described herein. A functional analog of a TDD is a molecule having the same or substantially the same biological function as said TDD (i.e., cytidine deaminase function). For example, the functional analog may be an isoform or a variant of the TDD, e.g., containing a portion of the TDD with or without additional amino acid residues and/or containing mutations relative to the TDD (e.g., a variant with at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to the TDD (e.g., a TDD comprising the amino acid sequence of any one of SEQ ID NOs: 72, 86-91, and 117-129) or its toxic domain (e.g., a toxic domain comprising the amino acid sequence of SEQ ID NO: 49, 81, 92, 95, 98, 101, 104, 107, 134, 143, 152, 157, 162, 167, 172, 177, 184, 189, 194, 199, 204, 209, 214, or 219)). In certain embodiments, the functional analogs are orthologs of a TDD described herein. In certain embodiments, a TDD ortholog may comprise an amino acid sequence at least 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to the amino acid sequence of said TDD (e.g., a TDD comprising the amino acid sequence of any one of SEQ ID NOs: 72, 86-91, and 117-129). In certain embodiments, a TDD ortholog may comprise a toxic domain with an amino acid sequence that is at least 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to the amino acid sequence of the toxic domain of a TDD described herein (e.g., a toxic domain comprising the amino acid sequence of SEQ ID NO: 49, 81, 92, 95, 98, 101, 104, 107, 134, 143, 152, 157, 162, 167, 172, 177, 184, 189, 194, 199, 204, 209, 214, or 219).

[0065] The term "percent identical" in the context of amino acid or nucleotide sequences refers to the percent of residues in two sequences that are the same when aligned for maximum correspondence. The percent identity of two sequences may be obtained by, e.g., BLAST® using default parameters (available at the U.S. National Library of Medicine's National Center for Biotechnology Information website). In some embodiments, the length of a reference sequence aligned for comparison purposes is at least 30%, (e.g., at least 40, 50, 60, 70, 80, or 90%, or 100%) of the reference sequence.

[0066] In certain embodiments, a cytidine deaminase described herein may target a cytidine in an AC sequence, a TC sequence, a GC sequence, a CC sequence, an AAC sequence, a TAC sequence, a GAC sequence, a CAC sequence, an ATC sequence, a TTC sequence, a GTC sequence, a CTC sequence, an AGC sequence, a TGC sequence, a GGC sequence, a CGC sequence, an ACC sequence, a TCC sequence, a GCC sequence, a CCC sequence, or any combination thereof. In certain embodiments, a cytidine deaminase described herein has increased efficiency and/or activity compared to DddA. In some

embodiments, the increased efficiency or activity may be, e.g., at any one or combination of the above target sequences.

[0067] It is also contemplated that adenine deaminases (e.g., TadA) may be used in the fusion proteins and cell editing systems described herein for conversion of A:T base pairs to G:C base pairs. In certain embodiments, a TDD may be mutated at residues that form the nucleotide pocket (e.g., a residue or combination of residues as described above for DddA) to allow the enzyme to act as an adenine deaminase, and/or to reduce TC sequence bias within the base editing window.

B. Zinc Finger Protein Domains

[0068] The fusion proteins described herein (such as ZFP-cytidine deaminase (e.g., ZFP-TDD), ZFP-cytidine deaminase inhibitor (e.g., ZFP-TDDI), or ZFP-nickase fusion proteins) comprise zinc finger protein (ZFP) domains. A "zinc finger protein" or "ZFP" refers to a protein having DNA-binding domains that are stabilized by zinc. ZFPs bind to DNA in a sequence-specific manner. The individual DNA-binding domains are referred to as "fingers." A ZFP has at least one finger, and each finger binds from two to four base pairs of nucleotides, typically three or four base pairs of DNA (contiguous or noncontiguous). Each zinc finger typically comprises approximately 30 amino acids and chelates zinc. An engineered ZFP 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 that bind the particular triplet or quadruplet sequence. *See, e.g.,* ZFP design methods described in detail in U.S. Pats. 5,789,538; 5,925,523; 6,007,988; 6,013,453; 6,140,081; 6,200,759; 6,453,242; 6,534,261; 6,979,539; and 8,586,526; and International Pat. Pubs. WO 95/19431; WO 96/06166; WO 98/53057; WO 98/53058; WO 98/53059; WO 98/53060; WO 98/54311; WO 00/27878; WO 01/60970; WO 01/88197; WO 02/016536; WO 02/099084; and WO 03/016496.

[0069] The ZFP domain of the present ZFP fusion proteins may include at least three (e.g., four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, or more) zinc fingers. Individual zinc fingers are typically spaced at three base pair intervals when bound to DNA. unless they are connected by engineered linkers capable of skipping one or more bases (*see*, e.g., Paschon et al., *Nat Commun.* (2019) 10:1133 and U.S. Pats. 8,772,453; 9,163,245;

9,394,531; and 9,982,245). A ZFP domain having three fingers typically recognizes a target site that includes 9 or 12 nucleotides. A ZFP domain having four fingers typically recognizes a target site that includes 12 to 15 nucleotides. A ZFP domain having five fingers typically recognizes a target site that includes 15 to 18 nucleotides. A ZFP domain having six fingers can recognize target sites that include 18 to 21 nucleotides.

[0070] The target specificity of the ZFP domain may be improved by mutations to the ZFP backbone as described in, e.g., U.S. Pat. Pub. 2018/0087072. The mutations include those made to residues in the ZFP backbone that can interact non-specifically with phosphates on the DNA backbone but are not involved in nucleotide target specificity. In some embodiments, these mutations comprise mutating a cationic amino acid residue to a neutral or anionic amino acid residue. In some embodiments, these mutations comprise mutating a polar amino acid residue to a neutral or non-polar amino acid residue. In further embodiments, mutations are made at positions (-4), (-5), (-9) and/or (-14) relative to the DNA-binding helix. In some embodiments, a zinc finger may comprise one or more mutations at positions (-4), (-5), (-9) and/or (-14). In further embodiments, one or more zinc fingers in a multi-finger ZFP domain may comprise mutations at positions (-4), (-5), (-9) and/or (-14). In some embodiments, the amino acids at positions (-4), (-5), (-9) and/or (-14) (e.g., an arginine (R) or lysine (K)) are mutated to an alanine (A), leucine (L), Ser (S), Asp (N), Glu (E), Tyr (Y), and/or glutamine (Q). In some embodiments, the R residue at position (-4) is mutated to Q.

[0071] Alternatively, 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. Pats. 5,420,032 and 6,833,252; Belfort et al., *Nucleic Acids Res.* (1997) 25:3379-88; Dujon et al., *Gene* (1989) 82:115-8; Perler et al., *Nucleic Acids Res.* (1994) 22:1125-7; Jasin, *Trends Genet.* (1996) 12:224-8; Gimble et al., *J Mol Biol.* (1996) 263:163-80; Argast et al., *J Mol Biol.* (1998) 280:345-53; 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., *Mol Cell* (2002) 10:895-905; Epinat et al., *Nucleic Acids Res.* (2003) 31:2952-62; Ashworth et al., *Nature* (2006) 441:656-59; Paques et al., *Current Gene Therapy* (2007) 7:49-66; and U.S. Pat. Pub. 2007/0117128.

[0072] In some embodiments, the present ZFP fusion proteins comprise one or more zinc finger domains. The domains may be linked together via an extendable flexible linker such

that, for example, one domain comprises one or more (e.g., 3, 4, 5, or 6) zinc fingers and another domain comprises additional one or more (e.g., 3, 4, 5, or 6) zinc fingers. In some embodiments, the linker is a standard inter-finger linker such that the finger array comprises one DNA-binding domain comprising 8, 9, 10, 11 or 12 or more fingers. In other embodiments, the linker is an atypical linker such as a flexible linker. For example, two ZFP domains may be linked to a cytidine deaminase, inhibitor, or nickase domain ("domain") such as those described herein in the configuration (from N terminus to C terminus) ZFP-ZFP-domain, domain-ZFP-ZFP, ZFP-domain-ZFP, or ZFP-domain-ZFP-domain (two ZFP-domain fusion proteins are fused together via a linker).

[0073] In some embodiments, the ZFP fusion proteins are "two-handed," i.e., they contain two zinc finger clusters (two ZFP domains) separated by intervening amino acids so that the two ZFP domains bind to two discontinuous target sites. An example of a two-handed type of zinc finger binding protein is SIP1, where a cluster of four zinc fingers is located at the amino terminus of the protein and a cluster of three fingers is located at the carboxyl terminus (*see* Remacle et al., *EMBO J.* (1999) 18(18):5073-84). Each cluster of zinc fingers in these proteins is able to bind to a unique target sequence and the spacing between the two target sequences can comprise many nucleotides.

[0074] The DNA-binding ZFP domains of the ZFP fusion proteins described herein direct the proteins to DNA target regions. In some embodiments, the DNA target region is at least 8 bps in length. For example, the target region may be 8 bps to 40 bps in length, such as 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, or 36 bps in length.

[0075] 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.

C. Base Editor Inhibitors

[0076] In some embodiments, the base editor systems described herein may include an inhibitor of the editor to better regulate temporally and spatially the base editing activity of the systems. For example, where the cytidine deaminase is a TDD as described herein, the inhibitor may be a TDDI that inhibits said TDD. Where the editor is the cytidine deaminase DddA, the inhibitor may be, e.g., DddI. In some embodiments, DddI has the amino acid sequence shown below.

MYADDFDGEI EIDEVDSLVE FLSRRPAFDA NNFVLTFEES GFPQLNIFAK
NDIAVVYYMD IGENFVSKGN SASGGTEKFY ENKLGGEVDL SKDCVVSKEQ
MIEAAKQFFA TKQRPEQLTW SEL (SEQ ID NO: 73)

[0077] Thus, in some embodiments, the base editor systems include a TDDI component in addition to ZFP-TDD fusion proteins. The TDDI component may be brought in close proximity to the TDD complex through a DNA-binding domain covalently fused to it, or through dimerization with a DNA-binding domain not covalently bound to it.

[0078] In some embodiments, the present base editing system comprises a ZFP-inhibitor fusion protein comprising a ZFP domain and an inhibitor domain, wherein the ZFP domain binds to a sequence in the DNA target region close (e.g., within 50-100 nt) to the ZFP-cytidine deaminase fusion proteins' binding sites. When this ZFP-inhibitor fusion protein is introduced to the cell, the inhibitor domain will be brought within close proximity to the cytidine deaminase complex and bind to the complex, thereby inhibiting the base editing activity of the cytidine deaminase at that locus. The presence of the sequence bound by the ZFP domain of ZFP-inhibitor determines the inhibitory activity of the inhibitor.

[0079] In some embodiments, the binding of the inhibitor domain to the cytidine deaminase complex may be regulated by an agent (e.g., a small molecule or a peptide). For example, the inhibitor domain may be fused to a dimerization domain, and its dimerization partner may be fused to a ZFP domain that binds to a sequence in the DNA target region close (e.g., within 50-100 nt) to the ZFP-cytidine deaminase fusion proteins' binding sites. The dimerization domains of the inhibitor and the ZFP may dimerize in the presence of a dimerization-inducing agent (e.g., a small molecule or peptide). In the presence of the agent, the inhibitor domain will be brought within close proximity to the DNA target region through dimerization, leading to binding and inactivation of the cytidine deaminase complex. Once the agent is withdrawn, the inhibitor domain will no longer be sequestered near the DNA target region and will detach from the cytidine deaminase complex, allowing the base editing process to proceed. Examples of such agents and dimerizing domains are shown in **Table 1** below:

Table 1. Dimerization Domains and Dimerization-Inducing Agents

Dimerization Partners		**Dimerizing Agent**
FKBP	FKBP	FK1012
FKBP	Calcineurin A (can)	FK506

Dimerization Partners		Dimerizing Agent
FKBP	CyP-Fas	FKCsA
FKBP	FRB (FKBP-rapamycin-binding) domain of mTOR	Rapamycin
GyrB	GyrB	Coumermycin
GAI	GID1 (gibberellin insensitive dwarf 1)	Gibberellin
ABI	PYL	Abscisic acid
ABI	PYR^{Mandi}	Mandipropamid
SNAP-tag	HaloTag	HaXS
eDHFR	HaloTag	TMp-HTag
Bcl-xL	Fab (AZ1)	ABT-737

[0080] Conversely, the dimerization of the domains fused to the ZFP and the inhibitor domains may be inhibited, rather than promoted, by a dimerization-inhibiting agent (e.g., a small molecule or peptide) such that the presence of the agent will permit activity of the cytidine deaminase complex. If the agent is withdrawn, the inhibitor domain will be able to bind to the cytidine deaminase complex, inhibiting the base editing process.

D. Uracil DNA Glycosylase Inhibitors

[0081] The term "uracil glycosylase inhibitor" or "UGI" as used herein, refers to a protein that can inhibit a uracil-DNA glycosylase base-excision repair enzyme. Upon detecting a G:U mismatch, the cell responds through base excision repair, initiated by excision of the mismatched uracil by uracil N-glycosylase (UNG). In some embodiments, a base editor system described herein further comprises one or more UGIs to protect the edited G:U intermediate from excision by UNG. In certain embodiments, a ZFP-cytidine deaminase (e.g., ZFP-TDD) fusion protein described herein may comprise one or more UGI domains, e.g., attached by a linker described herein. In some embodiments, the linker is an SGGS linker (SEQ ID NO: 245). The UGI domain(s) may be located at the N-terminus, the C-terminus, or any combination thereof, of the fusion protein (e.g., one UGI domain at the C-terminus, one UGI domain at the N-terminus, two UGI domains at the C-terminus, two UGI domains at the N-terminus, or any combination thereof). Additionally or alternatively, one or more UGI domains may be on a separate ZFP fusion protein ("ZFP-UGI"). In particular embodiments, the UGI domain comprises the amino acid sequence of SEQ ID NO: 20.

E. Nickases

[0082] In some embodiments, a base editor system described herein further comprises a nickase to create a single-stranded DNA break in the vicinity of the edited DNA target region (e.g., within 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nt from the edited base). The creation of the nick attracts DNA repair machinery such that the region downstream of the nick is excised and replaced, resulting in a fully edited double-stranded DNA target region. The nick may be, for example, 5' or 3' of the edited base on the same strand or the opposite strand.

[0083] In some embodiments, the base editor system described herein has a trimeric architecture to include nickase function. For example, one domain of a dimeric nickase may be fused to a ZFP-cytidine deaminase (e.g., a ZFP-TDD as described herein) and the other domain may be fused to an independent ZFP, such that binding of both ZFP domains to their DNA target regions results in an active nickase capable of producing a single-strand break. *See, e.g.,* **FIG. 9**.

[0084] In some embodiments, the base editor system described herein has a tetrameric architecture to include nickase function. In addition to the two ZFP-cytidine deaminase (e.g., ZFP-TDD as described herein) fusion proteins, such a system also comprises two ZFP-nickase proteins, wherein one domain of a dimeric nickase is fused to a first ZFP domain and the other domain fused to a second ZFP domain, such that binding of both ZFP domains to their DNA target regions results in an active nickase capable of producing a single-strand break.

[0085] In some embodiments, the nickase may be, for example, a ZFN nickase, a TALEN nickase, or a CRISPR/Cas nickase. In certain embodiments, the nickase is derived from a FokI DNA cleavage domain. In some embodiments, the FokI nickase comprises one or more mutations as compared to a parental FokI nickase, e.g., mutations to change the charge of the cleavage domain; mutations to residues that are predicted to be close to the DNA backbone based on molecular modeling and that show variation in FokI homologs; and/or mutations at other residues (*see, e.g.,* U.S. Pat. 8,623,618 and Guo et al., *J Mol Biol.* (2010) 400(1):96-107).

[0086] In the ZFP fusion proteins described herein, the nickase domain(s) may be positioned on either side of the DNA-binding ZFP domain, including at the N- or C-terminal side of the fusion molecule (N- and/or C-terminal to the ZFP domain). In some embodiments, a ZFP-cytidine deaminase (e.g., ZFP-TDD as described herein) fusion protein described herein comprises a cytidine deaminase domain at the N- or C- terminus and a nickase domain at the opposite terminus.

F. Peptide Linkers

[0087] In the fusion proteins described herein, the ZFP, cytidine deaminase (e.g., a TDD as described herein), inhibitor (e.g., a TDDI, such as DddI where the cytidine deaminase is DddA), nickase, and/or UGI domains may be positioned in any order relative to each other. In some embodiments, the domains may be associated with each other by direct peptidyl linkages, peptide linkers, or any combination thereof. In some embodiments, two or more of the domains may be associated with each other by dimerization (e.g., through a leucine zipper, a STAT protein N-terminal domain, or an FK506 binding protein).

[0088] In some embodiments, the ZFP, cytidine deaminase (e.g., a TDD as described herein), inhibitor (e.g., a TDDI, such as DddI where the cytidine deaminase is DddA), UGI, and/or nickase domains, and/or the zinc fingers within the ZFP domain, may be linked through a peptide linker, e.g., a noncleavable peptide linker of about 5 to 200 amino acids (e.g., 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 or more amino acids). Preferred linkers are typically flexible amino acid subsequences that are synthesized as a recombinant fusion protein. *See, e.g.,* U.S. Pats. 6,479,626; 6,903,185; 7,153,949; 8,772,453; and 9,163,245; and PCT Patent Pub. WO 2011/139349. The proteins described herein may include any combination of suitable linkers.

[0089] In some embodiments, the peptide linker is three to 30 amino acid residues in length and is rich in G and/or S. Non-limiting examples of such linkers are SGGS linkers (SEQ ID NO: 245) as well as G_4S-type linkers, i.e., linkers containing one or more (e.g., 2, 3, or 4) GGGGS (SEQ ID NO: 71) motifs, or variations of the motif (such as ones that have one, two, or three amino acid insertions, deletions, and substitutions from the motif).

[0090] In particular embodiments, a peptide linker used in a fusion protein described herein may be L0 (LRGSQLVKS; SEQ ID NO: 15), L7A (LRGSQLVKSKSEAAAR; SEQ ID NO: 16), L26 (LRGSQLVKSKSEAAARGGGGSGGGGS; SEQ ID NO: 17), L21 (LRGSQLVKSKSEAAARGGGGS; SEQ ID NO: 110), L18 (LRGSQLVKSKSEAAARGS; SEQ ID NO: 111), L13 (LRGSQLVKSKSGS; SEQ ID NO: 112), L11 (LRGSQLVKSGS; SEQ ID NO: 113), L9 (LRGSQLVGS; SEQ ID NO: 114), L6 (LRGSGS; SEQ ID NO: 115), or L4 (LRGS; SEQ ID NO: 116).

II. Base Editor Systems

[0091] The present disclosure provides base editor systems comprising the ZFP fusion proteins described herein. The base editor systems can be used to edit a cytosine base to a uracil base in a DNA target region, wherein the uracil is replaced by a thymine base during

DNA replication or repair. In certain embodiments, the editing results in the change of a targeted C:G base pair to a T:A base pair. **FIG. 1** illustrates a base editing system of the present disclosure.

[0092] Base editor systems as described herein can be used to knock out a gene (e.g., by changing a regular codon into a stop codon and/or by mutating a splice acceptor site to introduce exon skipping and/or frameshift mutations); introduce mutations into a control element of a gene (e.g., a promoter or enhancer region) to increase or reduce expression; correct disease-causing mutations (e.g., point mutations); and/or induce mutations that result in therapeutic benefits. The target DNA may be in a chromosome or in an extrachromosomal sequence (e.g., mitochondrial DNA) in a cell. The base editing may be performed *in vitro*, *ex vivo*, or *in vivo*.

[0093] In some embodiments, a base editor system described herein performs one or more codon conversions, e.g., CAA to TAA; CAG to TAG; CGA to TGA; or TGG to TAG, TGA, or TAA; or any combination thereof; thereby introducing stop codon(s).

[0094] The base editor systems of the present disclosure may comprise, in addition to ZFP-cytidine deaminase (e.g., ZFP-TDD as described herein) fusion proteins, components such as inhibitor domains (e.g., a TDDI, such as DddI where the cytidine deaminase is DddA), UGIs, and nickases, or any combination thereof, as described herein that may help regulate or improve the editing activity of the system. In certain embodiments, the system may be packaged within a single viral vector (e.g., an AAV vector).

[0095] In some embodiments, a base editor system of the present disclosure comprises a pair of ZFP-cytidine deaminase (e.g., ZFP-TDD as described herein) fusion proteins each comprising a cytidine deaminase half domain that lacks cytidine deaminase activity on its own, wherein binding of the ZFPs to their respective nucleotide targets results in an active cytidine deaminase molecule capable of editing a targeted C base to T (e.g., by replacing C with U, which is replaced by T during DNA replication or repair).

[0096] For example, in some embodiments, the base editor system may comprise: a) a first fusion protein (ZFP-TDD left) comprising: i) a first ZFP domain that binds to nucleotides of a double-stranded DNA target region on one side of the base targeted for editing; and ii) a TDD N-half domain; and b) a second fusion protein (ZFP-TDD right) comprising: i) a second ZFP domain that binds to nucleotides of the double-stranded DNA target region on the other side of the base targeted for editing; and ii) a TDD C-half domain; wherein binding of the ZFP-TDD left and the ZFP-TDD right to their respective nucleotides

results in an active TDD molecule capable of editing the DNA target region by changing the C base to T. The ZFP-TDDs and/or DNA target regions may be, e.g., as described herein.

[0097] In some embodiments, the base editor system may comprise: a) a first fusion protein (ZFP-TDDI) that binds to nucleotides within a first DNA target region, comprising: i) a zinc finger protein (ZFP) domain that binds to nucleotides within a first DNA target region; and

ii) a TDDI domain; b) a second fusion protein (ZFP-TDD left) comprising: i) a ZFP domain that binds to nucleotides of a second DNA target region on one side of the base targeted for editing; and ii) a TDD N-half domain; and c) a third fusion protein (ZFP-TDD right) comprising: i) a ZFP domain that binds to nucleotides of the second DNA target region on the other side of the base targeted for editing; and ii) a TDD C-half domain; wherein binding of ZFP-TDD left and ZFP-TDD right to their respective nucleotides results in an active TDD molecule capable of editing the second DNA target region by changing the C base to T; and wherein binding of ZFP-TDDI to the first DNA target region prevents editing of the second DNA target region by the TDD. The ZFP-TDDs, ZFP-TDDI, and DNA target regions may be, e.g., as described herein.

[0098] In some embodiments, the base editor system may comprise: a) a first fusion protein comprising: i) a zinc finger protein (ZFP) domain that binds to nucleotides within a first DNA target region, and ii) a dimerization domain; b) a second fusion protein comprising: i) a TDDI domain; and ii) a dimerization domain that partners with the dimerization domain of a); c) a third fusion protein (ZFP-TDD left) comprising: i) a ZFP domain that binds to nucleotides of a second DNA target region on one side of the base targeted for editing, and ii) a TDD N-half domain; and d) a fourth fusion protein (ZFP-TDD right) comprising: i) a ZFP domain that binds to nucleotides of the second DNA target region on the other side of the base targeted for editing, and ii) a TDD C-half domain; wherein binding of ZFP-TDD left and ZFP-TDD right to their respective nucleotides results in an active TDD molecule capable of editing the second DNA target region by changing the C base to T; and wherein dimerization of the fusion proteins of a) and b) to form ZFP-TDDI and binding of the ZFP of a) to the first DNA target region prevents editing of the second DNA target region by the TDD. The ZFP-TDDs, ZFP-TDDI, and/or DNA target regions may be, e.g., as described herein.

[0099] In some embodiments, the dimerization domains of the fusion proteins of a) and b) partner to form ZFP-TDDI in the presence of a dimerization-inducing agent, resulting in inhibition of TDD activity.

[00100] In some embodiments, the dimerization domains of the fusion proteins of a) and b) are inhibited from partnering to form ZFP-TDDI in the presence of a dimerizing-inhibiting agent, permitting TDD activity.

[00101] In some embodiments, the ZFP-TDDI is specific for a sequence to be protected from TDD base editing activity. For example, the ZFP domain may bind to an allele to be preserved in its unedited form (e.g., where another allele, such as a mutated allele, is targeted for editing), or a known site of off-target editing. In some embodiments, the TDD base editing may convert a regular codon into a stop codon in the unprotected allele.

[00102] In some embodiments, expression of ZFP-TDDI (or components thereof) may be under the control of an inducible promoter. In certain embodiments, such a system may be used as a "kill switch," wherein ZFP-TDDI protects an essential gene in a cell from being edited, and reducing or eliminating expression of ZFP-TDDI results in the death of the cell.

[00103] Where assembly of ZFP-TDDI is under the control of a dimerization-inducing or dimerization-inhibiting agent, base editing may be conditional upon the presence or absence of the agent. Such a conditional system may also be used for a "kill switch," e.g., wherein ZFP-TDDI protects an essential gene in a cell from being edited in the presence of a dimerization-inducing agent or in the absence of a dimerization-inhibiting agent, and removing or administering the agent, respectively, results in the death of the cell.

[00104] In certain embodiments, a base editor system of the present disclosure may be a multiplex system comprising more than one ZFP-TDD left and ZFP-TDD right pair; such a system may be capable of editing more than one DNA target region at a time. In particular embodiments, to increase editing specificity, the multiplex system comprises ZFP-TDD pairs wherein the TDD N-half and C-half domains are split at a different position in the TDD sequence (e.g., a position described herein) for each pair. In certain embodiments, the DNA target regions edited by the ZFP-TDD pairs of the multiplex system may be in different genes. In certain embodiments, the DNA target regions may be in the same gene.

[00105] In any of the above embodiments, the TDD and TDDI may be any described herein. In certain embodiments, the TDD may be DddA and the TDDI may be DddI. It is also contemplated that other cytidine deaminases and inhibitors may be used in place of the TDD and TDDI. In particular embodiments, a multiplex system described herein may comprise a first ZFP-cytidine deaminase pair and a second ZFP-cytidine deaminase pair, wherein the first and second pairs utilize different cytidine deaminases (e.g., selected from those described herein).

[00106] In some embodiments, the systems and methods described herein produce targeted editing of the DNA target region in at least 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% or 100% of the cells. In some embodiments, the edited cells exhibit little to no off-target indels (e.g., less than 5%, 4%, 3%, 2%, 1%, 0.5%, 0.2%, or 0.1% off-target indels). In some embodiments, the edited cells exhibit little to no off-target base editing (e.g., less than 5%, 4%, 3%, 2%, 1%, 0.5%, 0.2%, or 0.1% off-target base editing); however, as base editing of off-target sites may not be prone to translocations or other genomic arrangements, higher percentages may also be contemplated.

[00107] The present disclosure also provides nucleic acid molecules encoding the ZFP fusion proteins described herein, which may be part of a viral or non-viral vector. Further, the present disclosure provides a cell or population of cells comprising a base editor system as described herein, as well as descendants of such cells, wherein the cells comprise one or more edited bases.

III. Delivery of ZFP Fusion Proteins

[00108] A ZFP fusion protein of the present disclosure may be introduced to target cells as a protein, through a variety of methods (e.g., electroporation, fusion of the protein to a receptor ligand, lipid nanoparticles, cationic or anionic liposomes, or a nuclear localization signal (e.g., in combination with liposomes)). In other embodiments, the fusion protein is introduced to target cells through a nucleic acid molecule encoding it, for example, a DNA plasmid or mRNA. The nucleic acid molecule may be in a nucleic acid expression vector, which may include expression control sequences such as promoters, enhancers, transcription signal sequences, and transcription termination sequences that allow expression of the coding sequence for the ZFP fusion proteins.

[00109] In some embodiments, the promoter on the vector for directing ZFP fusion protein expression is a constitutively active promoter or an inducible promoter. Suitable promoters include, without limitation, a Rous sarcoma virus (RSV) long terminal repeat (LTR) promoter (optionally with an RSV enhancer), a cytomegalovirus (CMV) promoter (optionally with a CMV enhancer), a CMV immediate early promoter, a simian virus 40 (SV40) promoter, a dihydrofolate reductase (DHFR) promoter, a β-actin promoter, a phosphoglycerate kinase (PGK) promoter, an EFlα promoter, a Moloney murine leukemia virus (MoMLV) LTR, a creatine kinase-based (CK6) promoter, a transthyretin promoter (TTR), a thymidine kinase (TK) promoter, a tetracycline responsive promoter (TRE), a hepatitis B Virus (HBV)

promoter, a human α1-antitrypsin (hAAT) promoter, chimeric liver-specific promoters (LSPs), an E2 factor (E2F) promoter, the human telomerase reverse transcriptase (hTERT) promoter, a CMV enhancer/chicken β-actin/rabbit β-globin promoter (CAG promoter; Niwa et al., *Gene* (1991) 108(2):193-9), and an RU-486-responsive promoter. In addition, the promoter may include one or more self-regulating elements whereby the ZFP fusion protein can bind to and repress its own expression level to a preset threshold. *See* U.S. Pat. 9,624,498.

[00110] Any method of introducing the nucleotide sequence into a cell may be employed, including but not limited to, electroporation, calcium phosphate precipitation, microinjection, cationic or anionic liposomes, liposomes in combination with a nuclear localization signal, naturally occurring liposomes (e.g., exosomes), or viral transduction. In certain embodiments, the nucleotide sequence is in the form of mRNA and is delivered to a cell via electroporation.

[00111] For *in vivo* delivery of an expression vector, viral transduction may be used. A variety of viral vectors known in the art may be adapted by one of skill in the art for use in the present disclosure, for example, vaccinia vectors, adenoviral vectors, lentiviral vectors, poxyviral vectors, adeno-associated viral (AAV) vectors, retroviral vectors, and hybrid viral vectors. In some embodiments, the viral vector used herein is a recombinant AAV (rAAV) vector. Any suitable AAV serotype may be used. For example, the AAV may be AAV1, AAV2, AAV3, AAV3b, AAV4, AAV5, AAV6, AAV7, AAV8, AAV8.2, AAV9, AAV.PHP.B, AAV.PHP.eB, or AAVrh10, or of a novel serotype or a pseudotype such as AAV2/8, AAV2/5, AAV2/6, AAV2/9, or AAV2/6/9. In some embodiments, the expression vector is an AAV viral vector and is introduced to the target human cell by a recombinant AAV virion whose genome comprises the construct, including having the AAV Inverted Terminal Repeat (ITR) sequences on both ends to allow the production of the AAV virion in a production system such as an insect cell/baculovirus production system or a mammalian cell production system. The AAV may be engineered such that its capsid proteins have reduced immunogenicity or enhanced transduction ability in humans. Viral vectors described herein may be produced using methods known in the art. Any suitable permissive or packaging cell type may be employed to produce the viral particles. For example, mammalian (e.g., 293) or insect (e.g., sf9) cells may be used as the packaging cell line.

[00112] Any type of cell may be targeted for the base editing methods described herein. For example, the cells may be eukaryotic or prokaryotic. In some embodiments, the cells are mammalian (e.g., human) cells or plant cells. Human cells may can include, for example, T

cells, Natural Killer (NK) cells, NK T cells, alpha-beta T cells, gamma-delta T-cells, cytotoxic T lymphocytes (CTL), regulatory T cells, B cells, human embryonic stem cells, tumor-infiltrating lymphocytes (TIL) or a pluripotent stem cell from which lymphoid cells may be differentiated (e.g., an induced pluripotent stem cell (iPSC)). In some embodiments, the systems can be used to modify pluripotent stem cells prior to their differentiation into multiple cell types. For example, a lymphoid cell precursor may be modified prior to differentiation into lymphoid cell types such as regulatory T cells, effector T cells, natural killer cells, etc. The multiplex base editor systems of the present disclosure (comprising more than one ZFP-cytidine deaminase (e.g., ZFP-TDD) pair), in particular, can be used to prepare cells with multiple base edits at once, including pluripotent cells. In some embodiments, the multiplex systems may be used to prepare, e.g., allogeneic T cells. Where the systems comprise a ZFP-cytidine deaminase inhibitor (e.g., ZFP-TDDI) that can be induced to assemble in the presence or absence of a dimerization-regulating agent, as described herein, it is contemplated that the edited cells may be placed under the control of a "kill switch" activated upon administration of the agent.

[00113] For agricultural applications, any method for introduction of proteins or nucleic acid molecules to a plant cell is also contemplated, such as *Agrobacterium tumefaciens*-mediated T-DNA delivery.

IV. Pharmaceutical Applications

[00114] The present disclosure provides methods of editing a cytosine to a thymine base in cellular DNA, comprising delivering a base editor system described herein to a cell (e.g., from a patient), resulting in the replacement of a targeted C base with a T base. The cell may be within a patient (*in vivo* treatment), or a method as described herein may be performed on a cell removed from a patient and then the edited cell delivered to the patient (*ex vivo* treatment). In some embodiments, the cells are further manipulated *ex vivo* prior to use as a treatment. The term "treating" encompasses alleviation of symptoms, prevention of onset of symptoms, slowing of disease progression, improvement of quality of life, and increased survival. In some embodiments, a patient treated by the methods described herein is a mammal, e.g., a human.

[00115] In some embodiments, the methods of the present disclosure are used to edit a gene or regulatory sequence associated with a disease. For example, in certain embodiments, the base editing may correct a point mutation in a DNA sequence to restore normal gene expression or activity. In certain embodiments, the base editing may introduce a stop codon

into a deleterious gene (e.g., an oncogene). In certain embodiments, the base editing may introduce a mutation that results in a therapeutic benefit.

[00116] In some embodiments, the patient has cancer. In certain embodiments, the cell from the patient is further modified before or after base editing to provide resistance to a chemotherapeutic agent. The patient may then be treated with the chemotherapeutic agent, which in some embodiments may result in greater survival of edited over unedited cells.

[00117] In some embodiments, the patient has an autoimmune disorder.

[00118] In some embodiments, the patient has an autosomal dominant disease, such as autosomal dominant polycystic kidney disease.

[00119] In some embodiments, the patient has a mitochondrial disorder.

[00120] In some embodiments, the patient has sickle cell disease, hemophilia (e.g., hemophilia A, B, or C), cystic fibrosis, phenylketonuria, Tay-Sachs, prion disease, color blindness, a lysosomal storage disease (e.g., Fabry disease), Friedreich's ataxia, or prostate cancer.

[00121] In some embodiments, the methods of the present disclosure may target base editing to a particular allele of a gene, e.g., a wild-type or mutated allele. In certain embodiments, the allele may be associated with cancer. For example, the methods may target the V617F mutated allele of *JAK2*, which leads to constitutive tyrosine phosphorylation activity and plays a critical role in the expansion of myeloproliferative neoplasms. Knocking out expression of the allele with the V617F mutation, e.g., by introducing a stop codon, may facilitate successful treatment of *JAK2* V617F disorders.

[00122] The present disclosure further provides a pharmaceutical composition comprising elements of a base editor system described herein, such as a ZFP-cytidine deaminase (e.g., ZFP-TDD as described herein) pair and optionally a cytidine deaminase inhibitor (e.g., TDDI, such as DddI where the cytidine deaminase is DddA) component (e.g., a ZFP-cytidine deaminase inhibitor component), or nucleotide sequences encoding said elements (e.g., in viral or non-viral vectors as described herein). The pharmaceutical composition may further comprise a pharmaceutically acceptable carrier such as water, saline (e.g., phosphate-buffered saline), dextrose, glycerol, sucrose, lactose, gelatin, dextran, albumin, or pectin. In addition, the composition may contain auxiliary substances, such as, wetting or emulsifying agents, pH-buffering agents, stabilizing agents, or other reagents that enhance the effectiveness of the pharmaceutical composition. The pharmaceutical composition may contain delivery vehicles such as liposomes, nanocapsules, microparticles, microspheres, lipid particles, and vesicles.

[00123] In some embodiments, the base editor systems described herein can be engineered to target to a genomic locus chosen from 2B4 (CD244), 4-1BB (CD137), A2aR, AAVS1, ACTB, AID, ALB, B2M, B7.1, B7.2, B7-H2, B7-H3, B7-H4, B7-H6, BAFFR, BCL11A, BLAME (SLAMF8), BTLA, butyrophilins, CIITA, CCR5, CD100 (SEMA4D), CD103, CD3zeta, CD4, CD5, CD7, CD11a, CD11b, CD11c, CD11d, CD150, IPO-3), CD160, CD160 (BY55), CD18, CD19, CD2, CD27, CD28, CD29, CD30, CD4, CD40, CD47, CD48, CD49a, CD49D, CD49f, CD52, CD69, CD7, CD83, CD84, CD8alpha, CD8beta, CD96 (Tactile), CDS, CEACAM1, CISH, CRTAM, CTLA4, CXCR4, DCK, DGK, DGKA, DGKB, DGKD, DGKE, DGKG, DGKI, DGKK, DGKQ, DGKZ, DHFR, DNAM1 (CD226), EP2/4 receptors, adenosine receptors including A2AR, FAS, FASLG, GADS, GITR, GM-CSF, gp49B, HHLA2, HLA-A, HLA-B, HLA-C, HLA-DPA1, HLA-DPB1, HIV-LTR (long terminal repeat), HLA-DQA1, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-I, HVEM, HVEM, IA4, ICAM-1, ICOS, ICOS, ICOS (CD278), IFN-alpha/beta/gamma, IL-1 beta, IL-12, IL-15, IL-18, IL-23, IL2R beta, IL2R gamma, IL2RA, IL-6, IL7R alpha, ILT-2, ILT-4, immunoglobulin heavy chain loci, immunoglobulin light chain loci, ITGA4, ITGA4, ITGA6, ITGAD, ITGAE, ITGAL, ITGAM, ITGAX, ITGB1, ITGB2, ITGB7, KIR family receptors, KLRG1, Lag-3, LAIR-1, LAT, LIGHT, LTBR, Ly9 (CD229), MNK1/2, NKG2C, NKG2D, NKp30, NKp44, NKp46, NKp80 (KLRF1), OX2R, OX40, PAG/Cbp, PD-1, PD-L1, PD-L2, PGE2 receptors, PIR-B, PPP1R12C, PRNP1, PSGL1, PTPN2, RANCE/RANKL, RFX5, ROSA26, SELPLG (CD162), SIRPalpha (CD47), SLAM (SLAMF1, SLAMF4 (CD244, 2B4), SLAMF5, SLAMF6 (NTB-A, Ly108), SLAMF7, SLP-76, SOCS1, SOCS3, Tetherin, TGFBR2, TIGIT, TIM-1, TIM-3, TIM-4, TMIGD2, TRA, TRAC, TRB, TRD, TRG, TNF, TNF-alpha, TNFR2, TRIM5, TUBA1, VISTA, VLA1, or VLA-6.

[00124] It is understood that the ZFP fusion proteins and base editor systems described herein may be used in a method of treatment described herein, may be for use in a treatment described herein, or may be used in the manufacture of a medicament for a treatment described herein.

V. Agricultural Applications

[00125] The described systems and methods of editing a cytosine to a thymine base in cellular DNA may also be used in agricultural applications. For example, in certain embodiments, the base editing may correct one or more point mutations in a DNA sequence to restore normal gene expression or activity. In certain embodiments, the base editing may introduce a stop codon into one or more deleterious genes. In certain embodiments, the base

editing may introduce one or more beneficial mutations. In particular embodiments, the systems and methods described herein are used to edit a crop plant.

[00126] Unless otherwise defined herein, scientific and technical terms used in connection with the present disclosure shall have the meanings that are commonly understood by those of ordinary skill in the art. Exemplary methods and materials are described below, although methods and materials similar or equivalent to those described herein can also be used in the practice or testing of the present disclosure. In case of conflict, the present specification, including definitions, will control. Generally, nomenclature used in connection with, and techniques of, cardiology, medicine, medicinal and pharmaceutical chemistry, and cell biology described herein are those well-known and commonly used in the art. Enzymatic reactions and purification techniques are performed according to manufacturer's specifications, as commonly accomplished in the art or as described herein. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. Throughout this specification and embodiments, the words "have" and "comprise," or variations such as "has," "having," "comprises," or "comprising," will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers. It should also be noted that the term "or" is generally employed in its sense including "and/or" unless the content clearly dictates otherwise. As used herein the term "about" refers to a numerical range that is 10%, 5%, or 1% plus or minus from a stated numerical value within the context of the particular usage. Further, headings provided herein are for convenience only and do not interpret the scope or meaning of the claimed embodiments.

[00127] All publications and other references mentioned herein are incorporated by reference in their entirety. Although a number of documents are cited herein, this citation does not constitute an admission that any of these documents forms part of the common general knowledge in the art.

[00128] In order that this invention may be better understood, the following examples are set forth. These examples are for purposes of illustration only and are not to be construed as limiting the scope of the invention in any manner.

EXAMPLES

Example 1: ZFP-TDD Design

[00129] To prepare ZFP-DddA fusion protein pairs, the DddA peptide was split into two

halves (each lacking cytidine deaminase activity) at residue G1333, as described in Mok et al., *supra* ("DddA-G1333"), as well as at residues G1404 ("DddA-G1404") and G1407 ("DddA-G1407"). Eight left ZFPs and five right ZFPs were designed to target the DddA halves to a site at the human *CCR5* locus, such that the halves could dimerize at the target site and restore the catalytic activity of DddA. The left and right ZFP pairs cover a broad variety of different base editing windows from 2-bp to 24-bp **(FIG. 2A)**.

[00130] The N-terminal half of each split DddA pair was fused to the C-terminus of a left ZFP and the C-terminal half was fused to the C-terminus of a right ZFP, and vice-versa. For DddA-G1333, one of three different linkers (L0, L7A and L26) was used, whereas for DddA-G1404 and DddA-G1407, the L26 linker was used. For all other experiments, unless otherwise indicated, the L26 linker was used. A UGI (uracil DNA glycosylase inhibitor) domain was also fused to the C-terminus of each N-terminal and C-terminal half. All ZFP-DddA fusion constructs further contained a 3xFLAG tag as well as an SV40 nuclear localization signal fused to the N-terminus of the ZFP. An example of a left and right ZFP pair is shown in **FIG. 2B**.

[00131] The above-described sequences and the sequences of several prepared constructs are shown in **Table 2** below. Finger sequences are underlined and bolded in Left ZFPs #1-8 and Right ZFPs #1-5. The ZFPs in **Table 2** target the *CCR5* locus.

Table 2. Sequences of ZFP-DddA Components and Constructs (*CCR5* Locus ZFPs)

SEQ	Description	Sequence
1	3xFlag+NLS	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMA
2	Left ZFP #1	ERPFQCRICMRNFS**RSDSLSV**HIRTHTGEKPFACDICGRKFA**QSGS** **LTR**HTKIHTGSQKPFQCRICMRNFS**TSGHLSR**HIRTHTGEKPFACD ICGRKFA**QSGDLTR**HTKIHTHPRAPIPKPFQCRICMRNFS**MVCCRT** **L**HIRTHTGEKPFACDICGRKFA**RSANLTR**HTKIH
3	Left ZFP #2	ERPFQCRICMRNFS**RPYTLRL**HIRTHTGEKPFACDICGRKFA**RKYY** **LAK**HTKIHTGSQKPFQCRICMRNFS**DDWNLSQ**HIRTHTGEKPFACD ICGRKFA**RSANLTR**HTKIHTGEKPFQCRICMRKFA**QSAHRIT**HTKI H
4	Left ZFP #3	ERPFQCRICMRNFS**QSGALAR**HIRTHTGEKPFACDICGRKFA**LKQH** **LTR**HTKIHTGSQKPFQCRICMRNFS**QSGDLTR**HIRTHTGEKPFACD ICGRKFA**QSSDLRR**HTKIHTGSQKPFQCRICMRNFS**QSAHRKN**HIR THTGEKPFACDICGRKFA**RSAVRKN**HTKIH
5	Left ZFP #4	ERPFQCRICMRNFS**QSGALAR**HIRTHTGEKPFACDICGRKFA**LKQH** **LTR**HTKIHTGSQKPFQCRICMRNFS**QSGDLTR**HIRTHTGEKPFACD ICGRKFA**QSSDLRR**HTKIHTHPRAPIPKPFQCRICMRNFS**RSANLA** **R**HIRTHTGEKPFACDICGRKFA**TNQNRIT**HTKIH

SEQ	Description	Sequence
6	Left ZFP #5	ERPFQCRICMRNFS**RSDHLSA**HIRTHTGEKPFACDICGRKFA**CRRN LRN**HTKIHTGSQKPFQCRICMRNFS**MVCCRTL**HIRTHTGEKPFACD ICGRKFA**RSANLTR**HTKIHTGSQKPFQCRICMRNFS**TSSNRKT**HIR THTGEKPFACDICGRKFA**QSGHLSR**HTKIH
7	Left ZFP #6	ERPFQCRICMRNFS**DDWNLSQ**HIRTHTGEKPFACDICGRKFA**RSAN LTR**HTKIHTGSQKPFQCRICMRKFA**QSAHRIT**HTKIHTGEKPFQCR ICMRNFS**QSANRTT**HIRTHTGEKPFACDICGRKFA**QNAHRKT**HTKI H
8	Left ZFP #7	ERPFQCRICMRNFS**QSGDLTR**HIRTHTGEKPFACDICGRKFA**QSSD LRR**HTKIHTGSQKPFQCRICMRNFS**QSAHRKN**HIRTHTGEKPFACD ICGRKFA**RSAVRKN**HTKIHTGSQKPFQCRICMRNFS**QSANRTT**HIR THTGEKPFACDICGRKFA**RKYYLAK**HTKIH
9	Left ZFP #8	ERPFQCRICMRNFS**QSGDLTR**HIRTHTGEKPFACDICGRKFA**QSSD LRR**HTKIHTHPRAPIPKPFQCRICMRNFS**RSANLAR**HIRTHTGEKP FACDICGRKFA**TNQNRIT**HTKIHTGSQKPFQCRICMRNFS**QSGDLT R**HIRTHTGEKPFACDICGRKFA**RKDPLKE**HTKIH
10	Right ZFP #1	ERPFQCRICMRKFA**QSGNRTT**HTKIHTGEKPFQCRICMRNFS**TSSN RKT**HIRTHTGEKPFACDICGRKFA**AQWTRAC**HTKIHTGSQKPFQCR ICMRNFS**LRHHLTR**HIRTHTGEKPFACDICGRKFA**DRTGLRS**HTKI H
11	Right ZFP #2	ERPFQCRICMRNFS**QSGHLAR**HIRTHTGEKPFACDICGRKFA**NRHD RAK**HTKIHTPNPHRRTDPSHKPFQCRICMRNFS**QSADRTK**HIRTHT GEKPFACDICGRKFA**QSGSLTR**HTKIHTHPRAPIPKPFQCRICMRN FS**DRSTRIT**HIRTHTGEKPFACDICGRKFA**QNATRIN**HTKIH
12	Right ZFP #3	ERPFQCRICMRNFS**QSGHLAR**HIRTHTGEKPFACDICGRKFA**NRHD RAK**HTKIHTHPRAPIPKPFQCRICMRKFA**QSGNRTT**HTKIHTGEKP FQCRICMRNFS**TSSNRKT**HIRTHTGEKPFACDICGRKFA**AQWTRAC** HTKIH
13	Right ZFP #4	ERPFQCRICMRNFS**DIGYRAA**HIRTHTGEKPFACDICGRKFA**QSGN LAR**HTKIHTHPRAPIPKPFQCRICMRNFS**QSGHLAR**HIRTHTGEKP FACDICGRKFA**NRHDRAK**HTKIHTPNPHRRTDPSHKPFQCRICMRN FS**QSADRTK**HIRTHTGEKPFACDICGRKFA**QSGSLTR**HTKIH
14	Right ZFP #5	ERPFQCRICMRNFS**DRSNLSR**HIRTHTGEKPFACDICGRKFA**QSGD LTR**HTKIHTGSQKPFQCRICMRNFS**DIGYRAA**HIRTHTGEKPFACD ICGRKFA**QSGNLAR**HTKIHTHPRAPIPKPFQCRICMRNFS**QSGHLA R**HIRTHTGEKPFACDICGRKFA**NRHDRAK**HTKIH
15	L0	LRGSQLVKS
16	L7A	LRGSQLVKSKSEAAAR
17	L26	LRGSQLVKSKSEAAARGGGGSGGGGS
18	G1333-N	GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGG
19	G1333-C	PTPYPNYANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNM TETLLPENAKMTVVPPEGAIPVKRGATGETKVFTGNSNSPKSPTKG GC
82	G1404-N	GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGGPT

SEQ	Description	Sequence
		PYPNYANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMTE TLLPENAKMTVVPPEGAIPVKRG
83	G1404-C	ATGETKVFTGNSNSPKSPTKGGC
84	G1407-N	GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGGPT PYPNYANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMTE TLLPENAKMTVVPPEGAIPVKRGATG
85	G1407-C	ETKVFTGNSNSPKSPTKGGC
20	UGI	TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAY DESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML
21	G1333-N + UGI	GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGGSG GSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHT AYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML
22	G1333-C + UGI	PTPYPNYANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNM TETLLPENAKMTVVPPEGAIPVKRGATGETKVFTGNSNSPKSPTKG GCSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDI LVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML
23	Left ZFP #4_L0_G1333-N (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMAERPF QCRICMRNFSQSGALARHIRTHTGEKPFACDICGRKFALKQHLTRH TKIHTGSQKPFQCRICMRNFSQSGDLTRHIRTHTGEKPFACDICGR KFAQSSDLRRHTKIHTHPRAPIPKPFQCRICMRNFSRSANLARHIR THTGEKPFACDICGRKFATNQNRITHTKIHLRGSQLVKSGSYALGP YQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGGSGGSTNLSD IIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTD ENVMLLTSDAPEYKPWALVIQDSNGENKIKML*
24	Left ZFP #1_L7A_G1333-N (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMAERPF QCRICMRNFSRSDSLSVHIRTHTGEKPFACDICGRKFAQSGSLTRH TKIHTGSQKPFQCRICMRNFSTSGHLSRHIRTHTGEKPFACDICGR KFAQSGDLTRHTKIHTHPRAPIPKPFQCRICMRNFSMVCCRTLHIR THTGEKPFACDICGRKFARSANLTRHTKIHLRGSQLVKSKSEAAAR GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGGSG GSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHT AYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML*
25	Left ZFP #5_L7A_G1333-N (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMAERPF QCRICMRNFSRSDHLSAHIRTHTGEKPFACDICGRKFACRRNLRNH TKIHTGSQKPFQCRICMRNFSMVCCRTLHIRTHTGEKPFACDICGR KFARSANLTRHTKIHTGSQKPFQCRICMRNFSTSSNRKTHIRTHTG EKPFACDICGRKFAQSGHLSRHTKIHLRGSQLVKSKSEAAARGSYA LGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGGSGGSTN LSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDE STDENVMLLTSDAPEYKPWALVIQDSNGENKIKML*
26	Left ZFP #8_L7A_G1333-N (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMAERPF QCRICMRNFSQSGDLTRHIRTHTGEKPFACDICGRKFAQSSDLRRH TKIHTHPRAPIPKPFQCRICMRNFSRSANLARHIRTHTGEKPFACD ICGRKFATNQNRITHTKIHTGSQKPFQCRICMRNFSQSGDLTRHIR THTGEKPFACDICGRKFARKDPLKEHTKIHLRGSQLVKSKSEAAAR GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGGSG GSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHT AYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML*
27	Left ZFP #3_L7A_G1333-C (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMAERPF QCRICMRNFSQSGALARHIRTHTGEKPFACDICGRKFALKQHLTRH TKIHTGSQKPFQCRICMRNFSQSGDLTRHIRTHTGEKPFACDICGR KFAQSSDLRRHTKIHTGSQKPFQCRICMRNFSQSAHRKNHIRTHTG EKPFACDICGRKFARSAVRKNHTKIHLRGSQLVKSKSEAAARPTPY PNYANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMTETL

SEQ	Description	Sequence
		LPENAKMTVVPPEGAIPVKRGATGETKVFTGNSNSPKSPTKGGCSG GSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHT AYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML*
28	Left ZFP #1_L26_G1333-N (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMAERPF QCRICMRNFSRSDSLSVHIRTHTGEKPFACDICGRKFAQSGSLTRH TKIHTGSQKPFQCRICMRNFSTSGHLSRHIRTHTGEKPFACDICGR KFAQSGDLTRHTKIHTHPRAPIPKPFQCRICMRNFSMVCCRTLHIR THTGEKPFACDICGRKFARSANLTRHTKIHLRGSQLVKSKSEAAAR GGGGSGGGGSGSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLE SKVFSSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGN KPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKI KML*
29	Left ZFP #3_L26_G1333-N (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMAERPF QCRICMRNFSQSGALARHIRTHTGEKPFACDICGRKFALKQHLTRH TKIHTGSQKPFQCRICMRNFSQSGDLTRHIRTHTGEKPFACDICGR KFAQSSDLRRHTKIHTGSQKPFQCRICMRNFSQSAHRKNHIRTHTG EKPFACDICGRKFARSAVRKNHTKIHLRGSQLVKSKSEAAARGGGG SGGGGSGSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVF SSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPES DILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML*
30	Left ZFP #4_L26_G1333-N (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMAERPF QCRICMRNFSQSGALARHIRTHTGEKPFACDICGRKFALKQHLTRH TKIHTGSQKPFQCRICMRNFSQSGDLTRHIRTHTGEKPFACDICGR KFAQSSDLRRHTKIHTHPRAPIPKPFQCRICMRNFSRSANLARHIR THTGEKPFACDICGRKFATNQNRITHTKIHLRGSQLVKSKSEAAAR GGGGSGGGGSGSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLE SKVFSSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGN KPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKI KML*
31	Left ZFP #5_L26_G1333-N (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMAERPF QCRICMRNFSRSDHLSAHIRTHTGEKPFACDICGRKFACRRNLRNH TKIHTGSQKPFQCRICMRNFSMVCCRTLHIRTHTGEKPFACDICGR KFARSANLTRHTKIHTGSQKPFQCRICMRNFSTSSNRKTHIRTHTG EKPFACDICGRKFAQSGHLSRHTKIHLRGSQLVKSKSEAAARGGGG SGGGGSGSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVF SSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPES DILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML*
32	Left ZFP #8_L26_G1333-N (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMAERPF QCRICMRNFSQSGDLTRHIRTHTGEKPFACDICGRKFAQSSDLRRH TKIHTHPRAPIPKPFQCRICMRNFSRSANLARHIRTHTGEKPFACD ICGRKFATNQNRITHTKIHTGSQKPFQCRICMRNFSQSGDLTRHIR THTGEKPFACDICGRKFARKDPLKEHTKIHLRGSQLVKSKSEAAAR GGGGSGGGGSGSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLE SKVFSSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGN KPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKI KML*
33	Left ZFP #3_L26_G1333-C (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMAERPF QCRICMRNFSQSGALARHIRTHTGEKPFACDICGRKFALKQHLTRH TKIHTGSQKPFQCRICMRNFSQSGDLTRHIRTHTGEKPFACDICGR KFAQSSDLRRHTKIHTGSQKPFQCRICMRNFSQSAHRKNHIRTHTG EKPFACDICGRKFARSAVRKNHTKIHLRGSQLVKSKSEAAARGGGG SGGGGSPTPYPNYANAGHVEGQSALFMRDNGISEGLVFHNNPEGTC GFCVNMTETLLPENAKMTVVPPEGAIPVKRGATGETKVFTGNSNSP KSPTKGGCSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGN

SEQ	Description	Sequence
		KPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKI KML*
34	Left ZFP #4_L26_G1333-C (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMAERPF QCRICMRNFSQSGALARHIRTHTGEKPFACDICGRKFALKQHLTRH TKIHTGSQKPFQCRICMRNFSQSGDLTRHIRTHTGEKPFACDICGR KFAQSSDLRRHTKIHTHPRAPIPKPFQCRICMRNFSRSANLARHIR THTGEKPFACDICGRKFATNQNRITHTKIHLRGSQLVKSKSEAAAR GGGGSGGGGSPTPYPNYANAGHVEGQSALFMRDNGISEGLVFHNNP EGTCGFCVNMTETLLPENAKMTVVPPEGAIPVKRGATGETKVFTGN SNSPKSPTKGGCSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEE VIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNG ENKIKML*
35	Left ZFP #8_L26_G1333-C (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMAERPF QCRICMRNFSQSGDLTRHIRTHTGEKPFACDICGRKFAQSSDLRRH TKIHTHPRAPIPKPFQCRICMRNFSRSANLARHIRTHTGEKPFACD ICGRKFATNQNRITHTKIHTGSQKPFQCRICMRNFSQSGDLTRHIR THTGEKPFACDICGRKFARKDPLKEHTKIHLRGSQLVKSKSEAAAR GGGGSGGGGSPTPYPNYANAGHVEGQSALFMRDNGISEGLVFHNNP EGTCGFCVNMTETLLPENAKMTVVPPEGAIPVKRGATGETKVFTGN SNSPKSPTKGGCSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEE VIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNG ENKIKML*
36	Right ZFP #4_L0_G1333-C (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMAERPF QCRICMRNFSDIGYRAAHIRTHTGEKPFACDICGRKFAQSGNLARH TKIHTHPRAPIPKPFQCRICMRNFSQSGHLARHIRTHTGEKPFACD ICGRKFANRHDRAKHTKIHTPNPHRRTDPSHKPFQCRICMRNFSQS ADRTKHIRTHTGEKPFACDICGRKFAQSGSLTRHTKIHLRGSQLVK SPTPYPNYANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVN MTETLLPENAKMTVVPPEGAIPVKRGATGETKVFTGNSNSPKSPTK GGCSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESD ILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML*
37	Right ZFP #5_L0_G1333-C (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMAERPF QCRICMRNFSDRSNLSRHIRTHTGEKPFACDICGRKFAQSGDLTRH TKIHTGSQKPFQCRICMRNFSDIGYRAAHIRTHTGEKPFACDICGR KFAQSGNLARHTKIHTHPRAPIPKPFQCRICMRNFSQSGHLARHIR THTGEKPFACDICGRKFANRHDRAKHTKIHLRGSQLVKSPTPYPNY ANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMTETLLPE NAKMTVVPPEGAIPVKRGATGETKVFTGNSNSPKSPTKGGCSGGST NLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYD ESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML*
38	Right ZFP #1_L7A_G1333-C (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMAERPF QCRICMRKFAQSGNRTTHTKIHTGEKPFQCRICMRNFSTSSNRKTH IRTHTGEKPFACDICGRKFAAQWTRACHTKIHTGSQKPFQCRICMR NFSLRHHLTRHIRTHTGEKPFACDICGRKFADRTGLRSHTKIHLRG SQLVKSKSEAAARPTPYPNYANAGHVEGQSALFMRDNGISEGLVFH NNPEGTCGFCVNMTETLLPENAKMTVVPPEGAIPVKRGATGETKVF TGNSNSPKSPTKGGCSGGSTNLSDIIEKETGKQLVIQESILMLPEE VEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQD SNGENKIKML*
39	Right ZFP #5_L7A_G1333-C (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMAERPF QCRICMRNFSDRSNLSRHIRTHTGEKPFACDICGRKFAQSGDLTRH TKIHTGSQKPFQCRICMRNFSDIGYRAAHIRTHTGEKPFACDICGR KFAQSGNLARHTKIHTHPRAPIPKPFQCRICMRNFSQSGHLARHIR THTGEKPFACDICGRKFANRHDRAKHTKIHLRGSQLVKSKSEAAAR

SEQ	Description	Sequence
		PTPYPNYANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNM TETLLPENAKMTVVPPEGAIPVKRGATGETKVFTGNSNSPKSPTKG GCSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDI LVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML*
40	Right ZFP #4_L7A_G1333-C (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMAERPF QCRICMRNFSDIGYRAAHIRTHTGEKPFACDICGRKFAQSGNLARH TKIHTHPRAPIPKPFQCRICMRNFSQSGHLARHIRTHTGEKPFACD ICGRKFANRHDRAKHTKIHTPNPHRRTDPSHKPFQCRICMRNFSQS ADRTKHIRTHTGEKPFACDICGRKFAQSGSLTRHTKIHLRGSQLVK SKSEAAARPTPYPNYANAGHVEGQSALFMRDNGISEGLVFHNNPEG TCGFCVNMTETLLPENAKMTVVPPEGAIPVKRGATGETKVFTGNSN SPKSPTKGGCSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVI GNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGEN KIKML*
41	Right ZFP #1_L7A_G1333-N (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMAERPF QCRICMRKFAQSGNRTTHTKIHTGEKPFQCRICMRNFSTSSNRKTH IRTHTGEKPFACDICGRKFAAQWTRACHTKIHTGSQKPFQCRICMR NFSLRHHLTRHIRTHTGEKPFACDICGRKFADRTGLRSHTKIHLRG SQLVKSKSEAAARGSYALGPYQISAPQLPAYNGQTVGTFYYVNDAG GLESKVFSSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEV IGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGE NKIKML*
42	Right ZFP #1_L26_G1333-C (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMAERPF QCRICMRKFAQSGNRTTHTKIHTGEKPFQCRICMRNFSTSSNRKTH IRTHTGEKPFACDICGRKFAAQWTRACHTKIHTGSQKPFQCRICMR NFSLRHHLTRHIRTHTGEKPFACDICGRKFADRTGLRSHTKIHLRG SQLVKSKSEAAARGGGGSGGGGSPTPYPNYANAGHVEGQSALFMRD NGISEGLVFHNNPEGTCGFCVNMTETLLPENAKMTVVPPEGAIPVK RGATGETKVFTGNSNSPKSPTKGGCSGGSTNLSDIIEKETGKQLVI QESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPE YKPWALVIQDSNGENKIKML*
43	Right ZFP #5_L26_G1333-C (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMAERPF QCRICMRNFSDRSNLSRHIRTHTGEKPFACDICGRKFAQSGDLTRH TKIHTGSQKPFQCRICMRNFSDIGYRAAHIRTHTGEKPFACDICGR KFAQSGNLARHTKIHTHPRAPIPKPFQCRICMRNFSQSGHLARHIR THTGEKPFACDICGRKFANRHDRAKHTKIHLRGSQLVKSKSEAAAR GGGGSGGGGSPTPYPNYANAGHVEGQSALFMRDNGISEGLVFHNNP EGTCGFCVNMTETLLPENAKMTVVPPEGAIPVKRGATGETKVFTGN SNSPKSPTKGGCSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEE VIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNG ENKIKML*
44	Right ZFP #4_L26_G1333-C (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMAERPF QCRICMRNFSDIGYRAAHIRTHTGEKPFACDICGRKFAQSGNLARH TKIHTHPRAPIPKPFQCRICMRNFSQSGHLARHIRTHTGEKPFACD ICGRKFANRHDRAKHTKIHTPNPHRRTDPSHKPFQCRICMRNFSQS ADRTKHIRTHTGEKPFACDICGRKFAQSGSLTRHTKIHLRGSQLVK SKSEAAARGGGGSGGGGSPTPYPNYANAGHVEGQSALFMRDNGISE GLVFHNNPEGTCGFCVNMTETLLPENAKMTVVPPEGAIPVKRGATG ETKVFTGNSNSPKSPTKGGCSGGSTNLSDIIEKETGKQLVIQESIL MLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWA LVIQDSNGENKIKML*
45	Right ZFP #1_L26_G1333-N (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMAERPF QCRICMRKFAQSGNRTTHTKIHTGEKPFQCRICMRNFSTSSNRKTH IRTHTGEKPFACDICGRKFAAQWTRACHTKIHTGSQKPFQCRICMR

SEQ	Description	Sequence
		NFSLRHHLTRHIRTHTGEKPFACDICGRKFADRTGLRSHTKIHLRG SQLVKSKSEAAARGGGGSGGGGSGSYALGPYQISAPQLPAYNGQTV GTFYYVNDAGGLESKVFSSGGSGGSTNLSDIIEKETGKQLVIQESI LMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPW ALVIQDSNGENKIKML*
46	Right ZFP #5_L26_G1333-N (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMAERPF QCRICMRNFSDRSNLSRHIRTHTGEKPFACDICGRKFAQSGDLTRH TKIHTGSQKPFQCRICMRNFSDIGYRAAHIRTHTGEKPFACDICGR KFAQSGNLARHTKIHTHPRAPIPKPFQCRICMRNFSQSGHLARHIR THTGEKPFACDICGRKFANRHDRAKHTKIHLRGSQLVKSKSEAAAR GGGGSGGGGSGSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLE SKVFSSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGN KPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKI KML*

SEQ: SEQ ID NO.

Example 2: ZFP-DddA Base Editing in K562 Cells

[00132] To assay base editing in cells using same-linker ZFP-DddA pairs prepared according to the method described above, K562 (ATCC, CCL243) cells were obtained from the ATCC and were maintained in RPMI1640 with 10% FBS and 1× penicillin–streptomycin–glutamine (PSG) (Gibco, 10378-016) at 37 °C with 5% CO_2. 400 ng of pDNA encoding paired ZFP-DddA was electroporated into K562 cells using the SF cell line 96-well Nucleofector kit (Lonza, V4SC-2960) following the manufacturer's instructions. In brief, cells were washed twice with 1× PBS (divalent cation-free) and resuspended at 2×10^5 cells per 15 μL of supplemented SF cell line 96-well Nucleofector solution. For each transfection, 15 μL of the cell suspension was mixed with 5 μL of pDNA and transferred to the Lonza Nucleocuvette plate, then electroporated using the protocol for K562 cells (Nucleofector program 96-FF-120) on an Amaxa Nucleofector 96-well Shuttle System (Lonza). Electroporated cells were incubated at room temperature for 10 min and then transferred to 150 μL of prewarmed complete medium in a 96-well tissue culture plate. Cells were incubated for 72 h and then harvested for base editing quantification.

[00133] PCR primers for the *CCR5* locus were designed using Primer3 with the following optimal conditions: amplicon size of 200 nucleotides; a melting temperature of 60 °C; primer length of 20 nucleotides; and GC content of 50%. Sequences for the primers and amplicon are shown in **Table 3** below.

Table 3. *CCR5* Primer and Amplicon Sequences

SEQ	Description	Sequence
74	*CCR5* forward	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCAAGTGTGAT

SEQ	Description	Sequence
	primer	CACTTGGGTGG
75	*CCR5* reverse primer	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGATTCCCGAGTAGCAGATG
76	*CCR5* NGS amplicon	NNNNcaagtgtgatcacttgggtggtggctgtgtttgcgtctctcccaggaatcatctttaccagatctcaaaaagaaggtcttcattacacctgcagctctcattttccatacagtcagtatcaattctggaagaatttccagacattaaagatagtcatcttggggctggtcctgccgctgcttgtcatggtcatctgctactcgggaatcc

SEQ: SEQ ID NO.

[00134] Adaptors were added for a second PCR reaction to add the Illumina library sequences (forward primer: ACACGACGCTCTTCCGATCT (SEQ ID NO: 47); reverse primer: GACGTGTGCTCTTCCGAT (SEQ ID NO: 48)). The *CCR5* locus was amplified in 25 μL using 100 ng of genomic DNA with AccuPrime HiFi (Invitrogen). Primers were used at a final concentration of 0.1 μM with the following thermocycling conditions: initial melt of 95 °C for 5 min; 35 cycles of 95 °C for 30 s, 55 °C for 30 s and 68 °C for 40 s; and a final extension at 68 °C for 10 min. PCR products were diluted 1:20 in water. 2 μL of diluted PCR product was used in a 20 μL PCR reaction to add the Illumina library sequences with Phusion High-Fidelity PCR MasterMix with HF Buffer (NEB). Primers were used at a final concentration of 0.5 μM with the following conditions: initial melt of 98 °C for 30 s; 12 cycles of 98 °C for 10 s, 60 °C for 30 s and 72 °C for 40 s; and a final extension at 72 °C for 10 min. A second PCR reaction was then performed to add sample specific sequence barcodes. PCR libraries were purified using the QIAquick PCR purification kit (Qiagen). Samples were quantified with the Qubit dsDNA HS Assay kit (Invitrogen) and diluted to 2 nM. The libraries were then run according to the manufacturer's instructions on either an Illumina MiSeq using a standard 300-cycle kit or an Illumina NextSeq 500 using a mid-output 300-cycle kit. .

[00135] Results using DddA-G1333 are shown in **FIG. 3**. Base editing of >3% was achieved at all four positions in the *CCR5* base editing window (C9, C10, C18, and C24) with no noticeable indels. **FIG. 4** provides results for DddA-1397, DddA-G1404, and DddA-G1407 at positions C18 and C24. Notably, DddA-G1404 and DddA-G1407 showed increased efficiency and activity, particularly at C18. Base editing was not seen for any of the 17 GFP controls (data not shown).

Example 3: "Re-Wired" DddA Design

[00136] The DddA polypeptide chain was reconnected without performing standard circular

permutation by making residue 1398 the new N-terminus, linking the current C-terminus to residue 1334, linking residue 1397 to the current N-terminus, and making residue 1333 the new C-terminus, as shown below ("re-wired" DddA full):

>DddA full (residues 1290-1427 of SEQ ID NO: 72) (disordered residues italicized; 1333 and 1397 bolded):

*G*SYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSG**G**PTPYPNYANAGHVEGQSALFMRDN GISEGLVFHNNPEGTCGFCVNMTETLLPENAKMTVVPPE**G**AIPVKRGATGETKVFTGNSNSPKSP*TKG GC*
(SEQ ID NO: 49)

>re-wired DddA full (1398-C_term:1334-1397:linker (double underlined):N-term-1333, wherein single underlines indicate near junctions created by re-wiring):

AIPVKRGATGETKVFTGNSNSPKSP*TKGGC*PTPYPNYANAGHVEGQSALFMRDNGISEGLVFHNNPEG TCGFCVNMTETLLPENAKMTVVPPE**G**GGSGGS*G*SYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLE SKVFSSG**G** (SEQ ID NO: 50)

[00137] Two different strategies were then identified to split the re-wired DddA into two halves to make a functional non-toxic base editor, re-wired_G1309 and re-wired_N1357:

>re-wired_G1309-N:
AIPVKRGATGETKVFTGNSNSPKSP*TKGGC*PTPYPNYANAGHVEGQSALFMRDNGISEGLVFHNNPEG TCGFCVNMTETLLPENAKMTVVPPE**G**GGSGGS*G*SYALGPYQISAPQLPAYNG (SEQ ID NO: 51)

>re-wired_G1309-C:
QTVGTFYYVNDAGGLESKVFSSG**G** (SEQ ID NO: 52)

>re-wired_N1357-N:
AIPVKRGATGETKVFTGNSNSPKSP*TKGGC*PTPYPNYANAGHVEGQSALFMRDN (SEQ ID NO: 53)

>re-wired_N1357-C:
GISEGLVFHNNPEGTCGFCVNMTETLLPENAKMTVVPPE**G**GGSGGS*G*SYALGPYQISAPQLPAYNGQT VGTFYYVNDAGGLESKVFSSG**G** (SEQ ID NO: 54)

[00138] Respective ZFP-DddA base editors for the *CCR5* locus then were designed based on these split re-wired DddA architectures. *See*, e.g., **Table 4**. It is contemplated that when tested in K562 cells according to the protocols described above, the re-wired ZFP-DddA pairs will be able to perform C to T base editing. Such re-wired pairs may increase the specificity of multiplex base editor applications, as only the left and right arm of each split pair can form functional DddA.

Table 4. Sequences of Re-Wired ZFP-DddA Constructs (*CCR5* Locus)

SEQ	Description	Sequence
55	Left ZFP#8_L26_rewired _G1309-N (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMA ERPFQCRICMRNFSQSGDLTRHIRTHTGEKPFACDICGRKFA QSSDLRRHTKIHTHPRAPIPKPFQCRICMRNFSRSANLARHI RTHTGEKPFACDICGRKFATNQNRITHTKIHTGSQKPFQCRI CMRNFSQSGDLTRHIRTHTGEKPFACDICGRKFARKDPLKEH TKIHLRGSQLVKSKSEAAARGGGGSGGGGSAIPVKRGATGET KVFTGNSNSPKSPTKGGCPTPYPNYANAGHVEGQSALFMRDN GISEGLVFHNNPEGTCGFCVNMTETLLPENAKMTVVPPEGGG SGGSGSYALGPYQISAPQLPAYNGSGGSTNLSDIIEKETGKQ LVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVML LTSDAPEYKPWALVIQDSNGENKIKML*
56	Left ZFP#4_L26_rewired _G1309-N (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMA ERPFQCRICMRNFSQSGALARHIRTHTGEKPFACDICGRKFA LKQHLTRHTKIHTGSQKPFQCRICMRNFSQSGDLTRHIRTHT GEKPFACDICGRKFAQSSDLRRHTKIHTHPRAPIPKPFQCRI CMRNFSRSANLARHIRTHTGEKPFACDICGRKFATNQNRITH TKIHLRGSQLVKSKSEAAARGGGGSGGGGSAIPVKRGATGET KVFTGNSNSPKSPTKGGCPTPYPNYANAGHVEGQSALFMRDN GISEGLVFHNNPEGTCGFCVNMTETLLPENAKMTVVPPEGGG SGGSGSYALGPYQISAPQLPAYNGSGGSTNLSDIIEKETGKQ LVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVML LTSDAPEYKPWALVIQDSNGENKIKML*
57	Right ZFP#5_L26_rewired _G1309-C (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMA ERPFQCRICMRNFSDRSNLSRHIRTHTGEKPFACDICGRKFA QSGDLTRHTKIHTGSQKPFQCRICMRNFSDIGYRAAHIRTHT GEKPFACDICGRKFAQSGNLARHTKIHTHPRAPIPKPFQCRI CMRNFSQSGHLARHIRTHTGEKPFACDICGRKFANRHDRAKH TKIHLRGSQLVKSKSEAAARGGGGSGGGGSQTVGTFYYVNDA GGLESKVFSSGGSGGSTNLSDIIEKETGKQLVIQESILMLPE EVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWA LVIQDSNGENKIKML*
58	Right ZFP#2_L26_rewired _G1309-C (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMA ERPFQCRICMRNFSQSGHLARHIRTHTGEKPFACDICGRKFA NRHDRAKHTKIHTPNPHRRTDPSHKPFQCRICMRNFSQSADR TKHIRTHTGEKPFACDICGRKFAQSGSLTRHTKIHTHPRAPI PKPFQCRICMRNFSDRSTRITHIRTHTGEKPFACDICGRKFA QNATRINHTKIHLRGSQLVKSKSEAAARGGGGSGGGGSQTVG TFYYVNDAGGLESKVFSSGGSGGSTNLSDIIEKETGKQLVIQ ESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSD APEYKPWALVIQDSNGENKIKML*
59	Left ZFP#8_L26_rewired _G1309-C (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMA ERPFQCRICMRNFSQSGDLTRHIRTHTGEKPFACDICGRKFA QSSDLRRHTKIHTHPRAPIPKPFQCRICMRNFSRSANLARHI RTHTGEKPFACDICGRKFATNQNRITHTKIHTGSQKPFQCRI CMRNFSQSGDLTRHIRTHTGEKPFACDICGRKFARKDPLKEH TKIHLRGSQLVKSKSEAAARGGGGSGGGGSQTVGTFYYVNDA GGLESKVFSSGGSGGSTNLSDIIEKETGKQLVIQESILMLPE EVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWA LVIQDSNGENKIKML*
60	Left ZFP#4_L26_rewired _G1309-C	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMA ERPFQCRICMRNFSQSGALARHIRTHTGEKPFACDICGRKFA LKQHLTRHTKIHTGSQKPFQCRICMRNFSQSGDLTRHIRTHT

SEQ	Description	Sequence
	(incl UGI)	GEKPFACDICGRKFAQSSDLRRHTKIHTHPRAPIPKPFQCRI CMRNFSRSANLARHIRTHTGEKPFACDICGRKFATNQNRITH TKIHLRGSQLVKSKSEAAARGGGGSGGGGSQTVGTFYYVNDA GGLESKVFSSGGSGGSTNLSDIIEKETGKQLVIQESILMLPE EVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWA LVIQDSNGENKIKML*
61	Right ZFP#5_L26_rewired _G1309-N (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMA ERPFQCRICMRNFSDRSNLSRHIRTHTGEKPFACDICGRKFA QSGDLTRHTKIHTGSQKPFQCRICMRNFSDIGYRAAHIRTHT GEKPFACDICGRKFAQSGNLARHTKIHTHPRAPIPKPFQCRI CMRNFSQSGHLARHIRTHTGEKPFACDICGRKFANRHDRAKH TKIHLRGSQLVKSKSEAAARGGGGSGGGGSAIPVKRGATGET KVFTGNSNSPKSPTKGGCPTPYPNYANAGHVEGQSALFMRDN GISEGLVFHNNPEGTCGFCVNMTETLLPENAKMTVVPPEGGG SGGSGSYALGPYQISAPQLPAYNGSGGSTNLSDIIEKETGKQ LVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVML LTSDAPEYKPWALVIQDSNGENKIKML*
62	Right ZFP#2_L26_rewired _G1309-N (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMA ERPFQCRICMRNFSQSGHLARHIRTHTGEKPFACDICGRKFA NRHDRAKHTKIHTPNPHRRTDPSHKPFQCRICMRNFSQSADR TKHIRTHTGEKPFACDICGRKFAQSGSLTRHTKIHTHPRAPI PKPFQCRICMRNFSDRSTRITHIRTHTGEKPFACDICGRKFA QNATRINHTKIHLRGSQLVKSKSEAAARGGGGSGGGGSAIPV KRGATGETKVFTGNSNSPKSPTKGGCPTPYPNYANAGHVEGQ SALFMRDNGISEGLVFHNNPEGTCGFCVNMTETLLPENAKMT VVPPEGGGSGGSGSYALGPYQISAPQLPAYNGSGGSTNLSDI IEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDE STDENVMLLTSDAPEYKPWALVIQDSNGENKIKML*
63	Left ZFP#8_L26_rewired _G1357-N (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMA ERPFQCRICMRNFSQSGDLTRHIRTHTGEKPFACDICGRKFA QSSDLRRHTKIHTHPRAPIPKPFQCRICMRNFSRSANLARHI RTHTGEKPFACDICGRKFATNQNRITHTKIHTGSQKPFQCRI CMRNFSQSGDLTRHIRTHTGEKPFACDICGRKFARKDPLKEH TKIHLRGSQLVKSKSEAAARGGGGSGGGGSAIPVKRGATGET KVFTGNSNSPKSPTKGGCPTPYPNYANAGHVEGQSALFMRDN SGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPES DILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKI KML*
64	Left ZFP#4_L26_rewired _G1357-N (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMA ERPFQCRICMRNFSQSGALARHIRTHTGEKPFACDICGRKFA LKQHLTRHTKIHTGSQKPFQCRICMRNFSQSGDLTRHIRTHT GEKPFACDICGRKFAQSSDLRRHTKIHTHPRAPIPKPFQCRI CMRNFSRSANLARHIRTHTGEKPFACDICGRKFATNQNRITH TKIHLRGSQLVKSKSEAAARGGGGSGGGGSAIPVKRGATGET KVFTGNSNSPKSPTKGGCPTPYPNYANAGHVEGQSALFMRDN SGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPES DILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKI KML*
65	Right ZFP#5_L26_rewired _G1357-C (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMA ERPFQCRICMRNFSDRSNLSRHIRTHTGEKPFACDICGRKFA QSGDLTRHTKIHTGSQKPFQCRICMRNFSDIGYRAAHIRTHT GEKPFACDICGRKFAQSGNLARHTKIHTHPRAPIPKPFQCRI CMRNFSQSGHLARHIRTHTGEKPFACDICGRKFANRHDRAKH TKIHLRGSQLVKSKSEAAARGGGGSGGGGSGISEGLVFHNNP

SEQ	Description	Sequence
		EGTCGFCVNMTETLLPENAKMTVVPPEGGGSGGSGSYALGPY QISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGGSGGSTN LSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHT AYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML*
66	Right ZFP#2_L26_rewired _G1357-C (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMA ERPFQCRICMRNFSQSGHLARHIRTHTGEKPFACDICGRKFA NRHDRAKHTKIHTPNPHRRTDPSHKPFQCRICMRNFSQSADR TKHIRTHTGEKPFACDICGRKFAQSGSLTRHTKIHTHPRAPI PKPFQCRICMRNFSDRSTRITHIRTHTGEKPFACDICGRKFA QNATRINHTKIHLRGSQLVKSKSEAAARGGGGSGGGGSGISE GLVFHNNPEGTCGFCVNMTETLLPENAKMTVVPPEGGGSGGS GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSS GGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKP ESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGEN KIKML*
67	Left ZFP#8_L26_rewired _G1357-C (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMA ERPFQCRICMRNFSQSGDLTRHIRTHTGEKPFACDICGRKFA QSSDLRRHTKIHTHPRAPIPKPFQCRICMRNFSRSANLARHI RTHTGEKPFACDICGRKFATNQNRITHTKIHTGSQKPFQCRI CMRNFSQSGDLTRHIRTHTGEKPFACDICGRKFARKDPLKEH TKIHLRGSQLVKSKSEAAARGGGGSGGGGSGISEGLVFHNNP EGTCGFCVNMTETLLPENAKMTVVPPEGGGSGGSGSYALGPY QISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGGSGGSTN LSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHT AYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML*
68	Left ZFP#4_L26_rewired _G1357-C (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMA ERPFQCRICMRNFSQSGALARHIRTHTGEKPFACDICGRKFA LKQHLTRHTKIHTGSQKPFQCRICMRNFSQSGDLTRHIRTHT GEKPFACDICGRKFAQSSDLRRHTKIHTHPRAPIPKPFQCRI CMRNFSRSANLARHIRTHTGEKPFACDICGRKFATNQNRITH TKIHLRGSQLVKSKSEAAARGGGGSGGGGSGISEGLVFHNNP EGTCGFCVNMTETLLPENAKMTVVPPEGGGSGGSGSYALGPY QISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGGSGGSTN LSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHT AYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML*
69	Right ZFP#5_L26_rewired _G1357-N (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMA ERPFQCRICMRNFSDRSNLSRHIRTHTGEKPFACDICGRKFA QSGDLTRHTKIHTGSQKPFQCRICMRNFSDIGYRAAHIRTHT GEKPFACDICGRKFAQSGNLARHTKIHTHPRAPIPKPFQCRI CMRNFSQSGHLARHIRTHTGEKPFACDICGRKFANRHDRAKH TKIHLRGSQLVKSKSEAAARGGGGSGGGGSAIPVKRGATGET KVFTGNSNSPKSPTKGGCPTPYPNYANAGHVEGQSALFMRDN SGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPES DILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKI KML*
70	Right ZFP#2_L26_rewired _G1357-N (incl UGI)	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMA ERPFQCRICMRNFSQSGHLARHIRTHTGEKPFACDICGRKFA NRHDRAKHTKIHTPNPHRRTDPSHKPFQCRICMRNFSQSADR TKHIRTHTGEKPFACDICGRKFAQSGSLTRHTKIHTHPRAPI PKPFQCRICMRNFSDRSTRITHIRTHTGEKPFACDICGRKFA QNATRINHTKIHLRGSQLVKSKSEAAARGGGGSGGGGSAIPV KRGATGETKVFTGNSNSPKSPTKGGCPTPYPNYANAGHVEGQ SALFMRDNSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEE VIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQ

SEQ	Description	Sequence
		DSNGENKIKML*

SED: SEQ ID NO.

Example 4: Reshaping of the ZFP-DddA Binding Pocket

[00139] DddA-derived cytosine base editors are restricted to C to T editing and have a strong preference for TC dinucleotides within the base editing window. Various residues were identified for saturation mutagenesis to relax these restrictions and to increase the efficiency and/or activity of the enzyme, including Y1307, T1311, S1331, V1346, H1366, N1367, N1368, P1369, E1370, G1371, T1372, F1375, V1392, P1394, P1395, I1399, P1400, V1401, K1402, A1405, and T1406. The mutations are numbered with respect to SEQ ID NO: 72. Based on structural alignments between DddA and other base editors, including adenine deaminases, it was determined that these residues form the nucleotide pocket. DddA variants with mutations at positions E1370, N1368, and Y1307 were tested in K562 cells according to the protocols described above, using the left and right ZFP pairs shown in **FIG. 5**.

[00140] As shown in **FIGS. 6A-6C**, certain residue changes gave rise to an increase in efficiency/activity. Further, some residue changes altered the activity window of the DddA enzyme; such alterations may increase the precision and specificity of DddA-based reagents. Y1307 and N1368 both appeared sensitive to changes, with some mutations altering the activity profile of Y1307 (e.g., an almost 20x increase in activity at C18 in certain cases, and ability to access C9 and C10). E1370 appeared less sensitive to changes, with certain mutations showing a beneficial effect (e.g., E1370H, in the context of "Left_ZFP#4-G1333-N : Right _ZFP#5-G1333 -C").

Example 5: Combined ZFP-TDD + Nickase Approach to Base Editing

[00141] The efficiency of base editors can be increased by nicking the unmodified DNA strand with a nickase. The unmodified DNA strand then is recognized as newly synthesized by the cell, and the natural DNA repair machinery repairs the nicked DNA strand using the modified strand as a template. The unmodified strand can be nicked using a FokI-derived ZFN or TALEN or a CRISPR/Cas-derived nickase. **FIGs. 7A** and **7B** demonstrate a ZFP-TDD base editing design and results, respectively, with a CRISPR/Cas9 nickase. However, all three approaches require the delivery of two additional constructs (two peptides for ZFN or TALEN nickases; one peptide and one sgRNA for CRISPR/Cas nickases; **FIG. 8**).

[00142] A trimeric ZFP-TDD base editor architecture was developed to overcome this limitation, facilitating delivery and also making it more likely that the base editing and DNA nicking will happen simultaneously, increasing editing efficiency. With such a trimeric architecture, one half of a dimeric FokI nickase may be fused to the N-terminus of the left or right ZFP-TDD and the corresponding other half of the FokI nickase may be targeted to the site of interest through an independent ZFP-FokI peptide (**FIG. 9**). Sequences for nickase experiments using DddA may be found in **Table 5** below, with the ZFP design shown in **FIG. 10** (Left_ZFP#4 + Right_ZFP#1 + Nickase ZFP #2, or Left_ZFP#4 + Right_ZFP#5 + Nickase ZFP #1).

Table 5. Sequences of ZFP-Nickase Constructs (*CCR5* Locus)

SEQ	Description	Sequence
77	FokI(ELD)-Right_ZFP#5-G1333-C	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMGQLVKSE LEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYG YRGKHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEM ERYVEENQTRDKHLNPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLT RLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINFSG AQGSTLDFRPFQCRICMRNFSDRSNLSRHIRTHTGEKPFACDICGRKF AQSGDLTRHTKIHTGSQKPFQCRICMRNFSDIGYRAAHIRTHTGEKPF ACDICGRKFAQSGNLARHTKIHTHPRAPIPKPFQCRICMRNFSQSGHL ARHIRTHTGEKPFACDICGRKFANRHDRAKHTKIHLRGSQLVKSKSEA AARGGGGSGGGGSPTPYPNYANAGHVEGQSALFMRDNGISEGLVFHNN PEGTCGFCVNMTETLLPENAKMTVVPPEGAIPVKRGATGETKVFTGNS NSPKSPTKGGCSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIG NKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIK ML*
78	Nickase #1 (ZFP-FokI (KKR_F450N))	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMGQLVKSE LEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYG YRGKHLGGSRKPNGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEM QRYVKENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLT RLNRKTNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINFSG AQGSTLDFRPFQCRICMRNFSCSNNLPTHIRTHTGEKPFACDICGRKF ADRSNLTRHTKIHTGSQKPFQCRICMRNFSTSGNLTRHIRTHTGEKPF ACDICGRKFAQAENLKSHTKIHTGEKPFQCRICMRKFADRSTLRQHTK IHLRQKD*
79	FokI(ELD)-Right_ZFP#1-G1333-N	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMGQLVKSE LEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYG YRGKHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEM ERYVEENQTRDKHLNPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLT RLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINFSG AQGSTLDFRPFQCRICMRKFAQSGNRTTHTKIHTGEKPFQCRICMRNF STSSNRKTHIRTHTGEKPFACDICGRKFAAQWTRACHTKIHTGSQKPF QCRICMRNFSLRHHLTRHIRTHTGEKPFACDICGRKFADRTGLRSHTK IHLRGSQLVKSKSEAAARGGGGSGGGGSGSYALGPYQISAPQLPAYNG QTVGTFYYVNDAGGLESKVFSSGGSGGSTNLSDIIEKETGKQLVIQES ILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWA LVIQDSNGENKIKML*

80	Nickase #2 (ZFP-FokI (KKR_F450N))	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMGQLVKSE LEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYG YRGKHLGGSRKPNGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEM QRYVKENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLT RLNRKTNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINFSG AQGSTLDFRPFQCRICMRKFARNADRKKHTKIHTGEKPFQCRICMRNF STSSNRKTHIRTHTGEKPFACDICGRKFAQSGHLSRHTKIHTHPRAPI PKPFQCRICMRNFSDRSALSRHIRTHTGEKPFACDICGRKFATSSNRK THTKIHLRQKD*

[00143] The trimeric ZFP-DddA-nickase system was tested in K562 cells according to the protocols described above. As shown in **FIG. 11**, the trimeric ZFP-DddA-nickase system demonstrated a higher level of base editing activity than CRISPR-based nickases, with around 70% base edits in some cases, and a lower level of indels that approached background. In addition to outperforming the CRISPR-based nickase system, the trimeric ZFP-TDD-nickase system may be highly advantageous in its compact size, which may fit into a single viral vector such as AAV, unlike other platforms such as CRISPR/Cas and TALE-TDD base editor systems.

Example 6: Base Editing Activity of TDDs in K562 Cells

[00144] 19 other potential cytidine deaminases were identified (**Table 6**) and were tested for base editing activity.

Table 6. TDD Information

No.	NCBI No.	SEQ	Organism
TDD1	WP_069977532.1	86	*Streptomyces rubrolavendulae*
TDD2	WP_021798742.1	87	*Propionibacterium acidifaciens*
TDD3	QNM04114	88	*Lachnospiraceae bacterium sunii NSJ-8*
TDD4	WP_181981612	89	*Ruminococcus bicirculans*
TDD5	AXI73669.1	90	*Streptomyces cavourensis*
TDD6	WP_195441564	91	*Roseburia intestinalis*
TDD7	AVT32940.1	117	*Plantactinospora sp BC1*
TDD8	WP_189594293.1	118	*Streptomyces massasporeous*
TDD9	TCP42004.1	119	*Streptomyces sp. BK438*
TDD10	WP_171906854.1	120	*Jiangella alba*
TDD11	WP_174422267.1	121	*Burkholderia diffusa*
TDD12	WP_059728184.1	122	*Burkholderia ubonensis*
TDD13	WP_133186147.1	123	*Paraburkholderia guartelaensis*
TDD14	WP_083941146.1	124	*Pseudoduganella violaceinigra*
TDD15	WP_082507154.1	125	*Duganella sp. Root336D2*
TDD16	WP_044236021.1	126	*Chondromyces apiculatus*
TDD17	WP_165374601.1	127	*Sorangium cellulosum*
TDD18	NLI59004.1	128	*Clostridium sp.*

TDD19	KAB8140648.1	129	*Chloroflexia bacterium SDU3*

SEQ: SEQ ID NO:

[00145] TDDs described above were substituted for DddA in the base editing systems described in the above Examples, and were tested in K562 cells according to the described protocols for base editing at a *CCR5* locus, using the *CCR5*-targeting ZFPs described above, and/or at a *CIITA* locus ("site 2"), using the *CIITA*-targeting ZFPs described below (*see* **Table 7**). Sequences for the *CIITA* primers and amplicon are shown in **Table 8** below.

Table 7. *CIITA* Site 2 Zinc Finger Proteins

SEQ	Description	Sequence
241	CIITA_site_2_left_6	ERPFQCRICMRNFSRSAHLSRHIRTHTGEKPFACDICGRKFAT SGHLSRHTKIHTHPRAPIPKPFQCRICMRNFSDSSHRTRHIRT HTGEKPFACDICGRKFAAKWNLDAHTKIHTGSQKPFQCRICMR NFSRPYTLRLHIRTHTGEKPFACDICGRKFALRHHLTRHTKIH
242	CIITA_site_2_right_1	ERPFQCRICMRNFSQSGHLARHIRTHTGEKPFACDICGRKFAR KWTLQGHTKIHTGSQKPFQCRICMRNFSIRSTLRDHIRTHTGE KPFACDICGRKFAHRSSLRRHTKIHTGSQKPFQCRICMRNFSQ SGNLARHIRTHTGEKPFACDICGRKFARNVDLIHHTKIH
243	CIITA_site_2_right_5	ERPFQCRICMRNFSIRSTLRDHIRTHTGEKPFACDICGRKFAH RSSLRRHTKIHTGSQKPFQCRICMRNFSQSGNLARHIRTHTGE KPFACDICGRKFARNVDLIHHTKIHTGSQKPFQCRICMRNFSR SDVLSEHIRTHTGEKPFACDICGRKFATSGHLSRHTKIH

SEQ: SEQ ID NO.

[00146] One member of each TDD split was fused to the C-terminus of a left ZFP, and the other member was fused to the C-terminus of a right ZFP, using the L26 linker (SEQ ID NO: 17). A UGI (uracil DNA glycosylase inhibitor) domain (SEQ ID NO: 20) was also fused to the C-terminus of each N-terminal and C-terminal half with an SGGS linker (SEQ ID NO: 245). All ZFP fusion constructs further contained a 3xFLAG tag as well as an SV40 nuclear localization signal (SEQ ID NO: 1) fused to the N-terminus of the ZFP.

Table 8. *CIITA* Site 2 Primer and Amplicon Sequences

SEQ	Description	Sequence
224	*CIITA* site 2 forward primer	ACACGACGCTCTTCCGATCTNNNNCTGGGGCAGCTGATCACATGT
225	*CIITA* site 2 reverse primer	GACGTGTGCTCTTCCGATCTCTTCCATCCCCTCCCCAAG
226	*CIITA* site 2 NGS amplicon	NNNNCTGGGGCAGCTGATCACATGTTTTCTCTGCAGCCTTCCCAGAG GAGCTTCCGGCAGACCTGAAGCACTGGAAGCCAGGTGTGCAGGGCAG GTGGGCTGGGGTTGGGAAGGGTGGATGCCTTGGGGAGGGGATGGAAG

SEQ: SEQ ID NO.

[00147] Sequences used for the TDDs are included in **Table 9** below. For certain TDDs, a variant toxic domain was also tested (indicated by "b" after the TDD indicator, e.g., "TDD2b" for TDD2).

Table 9. Sequences of TDD Toxic Domains and Splits

No.	Description	Sequence	SEQ
TDD1	toxic domain	VAGNRAFTQRARTYNLTVADLHTYYVLAGQTPVLVH NANCGPHLKDLQKDYPRRTVGILDVGTDQLPMISGPGG QSGLLKNLPGRTKANGEHVETHAAAFLRMNPGVRKAV LYIDYPTGTCGTCRSTLPDMLPEGVQLWVISPRRTEKFT GLPD	92
	G2278-N	VAGNRAFTQRARTYNLTVADLHTYYVLAGQTPVLVH NANCGPHLKDLQKDYPRRTVGILDVGTDQLPMISGPGG	93
	G2278-C	QSGLLKNLPGRTKANGEHVETHAAAFLRMNPGVRKAV LYIDYPTGTCGTCRSTLPDMLPEGVQLWVISPRRTEKFT GLPD	94
	S2346-N	VAGNRAFTQRARTYNLTVADLHTYYVLAGQTPVLVH NANCGPHLKDLQKDYPRRTVGILDVGTDQLPMISGPGG QSGLLKNLPGRTKANGEHVETHAAAFLRMNPGVRKAV LYIDYPTGTCGTCRSTLPDMLPEGVQLWVIS	130
	S2346-C	PRRTEKFTGLPD	131
TDD2	toxic domain	LSTTGKNVLGHFEPTPTTPQGTSSDTIAEMLNSASQPGR TAGVLDIDGELTPLTSGRPSLPNYIASGHVEGQAAMIM RQQQVQSATVYHDNPNGTCGYCYSQLPTLLPEGAALD VVPPAGTVPPSNRWHNGGPSFIGNSSEPKPWPR	95
	G1794-N	LSTTGKNVLGHFEPTPTTPQGTSSDTIAEMLNSASQPGR TAGVLDIDGELTPLTSG	96
	G1794-C	RPSLPNYIASGHVEGQAAMIMRQQQVQSATVYHDNPN GTCGYCYSQLPTLLPEGAALDVVPPAGTVPPSNRWHN GGPSFIGNSSEPKPWPR	97
	P1861-N	LSTTGKNVLGHFEPTPTTPQGTSSDTIAEMLNSASQPGR TAGVLDIDGELTPLTSGRPSLPNYIASGHVEGQAAMIM RQQQVQSATVYHDNPNGTCGYCYSQLPTLLPEGAALD VVPPAGTVP	132
	P1861-C	PSNRWHNGGPSFIGNSSEPKPWPR	133
TDD2b	toxic domain	PTPTTPQGTSSDTIAEMLNSASQPGRTAGVLDIDGELTP LTSGRPSLPNYIASGHVEGQAAMIMRQQQVQSATVYH DNPNGTCGYCYSQLPTLLPEGAALDVVPPAGTVPPSNR WHNGGPSFIGNSSEPKPWPR	134
	G1794-N	PTPTTPQGTSSDTIAEMLNSASQPGRTAGVLDIDGELTP LTSG	135
	G1794-C	RPSLPNYIASGHVEGQAAMIMRQQQVQSATVYHDNPN GTCGYCYSQLPTLLPEGAALDVVPPAGTVPPSNRWHN GGPSFIGNSSEPKPWPR	136
	P1861-N	PTPTTPQGTSSDTIAEMLNSASQPGRTAGVLDIDGELTP LTSGRPSLPNYIASGHVEGQAAMIMRQQQVQSATVYH DNPNGTCGYCYSQLPTLLPEGAALDVVPPAGTVP	137
	P1861-C	PSNRWHNGGPSFIGNSSEPKPWPR	138
TDD3	toxic domain	MSLPEYDGTTTHGVLVLDDGTQIGFTSGNGDPRYTNYR	98

		NNGHVEQKSALYMRENNISNATVYHNNTNGTCGYCN TMTATFLPEGATLTVVPPENAVANNSRAIDYVKTYTGT SNDPKISPRYKGN	
	G30-N	MSLPEYDGTTTHGVLVLDDGTQIGFTSGNG	99
	G30-C	DPRYTNYRNNGHVEQKSALYMRENNISNATVYHNNTN GTCGYCNTMTATFLPEGATLTVVPPENAVANNSRAIDY VKTYTGTSNDPKISPRYKGN	100
	N94-N	DPRYTNYRNNGHVEQKSALYMRENNISNATVYHNNTN GTCGYCNTMTATFLPEGATLTVVPPEN	139
	N94-C	AVANNSRAIDYVKTYTGTSNDPKISPRYKGN	140
TDD4	toxic domain	HTYHVGKCRLLVHNANCNQEKPVLPKYDGKTTEGVM VTPDGKQISFKSGNSSTPSYPQYKAQSASHVEGKAALY MRENGINEATVFHNNPNGTCGFCDRQVPALLPKGAKL TVVPPSNSVANNVRAIPVPKTYIGNSTVPKIK	101
	T161-N	HTYHVGKCRLLVHNANCNQEKPVLPKYDGKTTEGVM VTPDGKQISFKSGNSST	102
	T161-C	PSYPQYKAQSASHVEGKAALYMRENGINEATVFHNNP NGTCGFCDRQVPALLPKGAKLTVVPPSNSVANNVRAIP VPKTYIGNSTVPKIK	103
	A229-N	HTYHVGKCRLLVHNANCNQEKPVLPKYDGKTTEGVM VTPDGKQISFKSGNSSTPSYPQYKAQSASHVEGKAALY MRENGINEATVFHNNPNGTCGFCDRQVPALLPKGAKL TVVPPSNSVA	141
	A229-C	NNVRAIPVPKTYIGNSTVPKIK	142
TDD4b	toxic domain	ANCNQEKPVLPKYDGKTTEGVMVTPDGKQISFKSGNSS TPSYPQYKAQSASHVEGKAALYMRENGINEATVFHNN PNGTCGFCDRQVPALLPKGAKLTVVPPSNSVANNVRAI PVPKTYIGNSTVPKIK	143
	T161-N	ANCNQEKPVLPKYDGKTTEGVMVTPDGKQISFKSGNSS T	144
	T161-C	PSYPQYKAQSASHVEGKAALYMRENGINEATVFHNNP NGTCGFCDRQVPALLPKGAKLTVVPPSNSVANNVRAIP VPKTYIGNSTVPKIK	145
	A229-N	ANCNQEKPVLPKYDGKTTEGVMVTPDGKQISFKSGNSS TPSYPQYKAQSASHVEGKAALYMRENGINEATVFHNN PNGTCGFCDRQVPALLPKGAKLTVVPPSNSVA	146
	A229-C	NNVRAIPVPKTYIGNSTVPKIK	147
TDD5	toxic domain	VQITAIKRWTETATVHNLTVADLHTYYVLAGKTPVLV HNENCGPNLKDLPKDYDRRTVGILDVGTDQLPMISGPG GQSGLLKNLPGRTKANTDHVEAHTAAFLRMNPGIRKA VLYIDYPTGTCGTCGSTLPDMLPEGVQLWVISPRKTEK FAGLPD	104
	G2299-N	VQITAIKRWTETATVHNLTVADLHTYYVLAGKTPVLV HNENCGPNLKDLPKDYDRRTVGILDVGTDQLPMISGPG G	105
	G2299-C	QSGLLKNLPGRTKANTDHVEAHTAAFLRMNPGIRKAV LYIDYPTGTCGTCGSTLPDMLPEGVQLWVISPRKTEKF AGLPD	106
	S2367-N	VQITAIKRWTETATVHNLTVADLHTYYVLAGKTPVLV HNENCGPNLKDLPKDYDRRTVGILDVGTDQLPMISGPG GQSGLLKNLPGRTKANTDHVEAHTAAFLRMNPGIRKA VLYIDYPTGTCGTCGSTLPDMLPEGVQLWVIS	148
	S2367-C	PRKTEKFAGLPD	149
TDD6	toxic domain	SAGAGESGRKTISLPEYDGTTTHGVLVLDDGTQIGFTSG	107

		NGDPRYTNYRNNGHVEQKSALYMRENNISNATVYHN NTNGTCGYCNTMTATFLPEGATLTVVPPENAVANNSR AIDYVKTYTGTSNDPKISPRYKGN	
	N2313-N	SAGAGESGRKTISLPEYDGTTTHGVLVLDDGTQIGFTSG N	108
	N2313-C	GDPRYTNYRNNGHVEQKSALYMRENNISNATVYHNNT NGTCGYCNTMTATFLPEGATLTVVPPENAVANNSRAID YVKTYTGTSNDPKISPRYKGN	109
	R2385-N	SAGAGESGRKTISLPEYDGTTTHGVLVLDDGTQIGFTSG NGDPRYTNYRNNGHVEQKSALYMRENNISNATVYHN NTNGTCGYCNTMTATFLPEGATLTVVPPENAVANNSR	150
	R2385-C	AIDYVKTYTGTSNDPKISPRYKGN	151
TDD6b	toxic domain	DPSGYDSQYPCKEEMSAGAGESGRKTISLPEYDGTTTH GVLVLDDGTQIGFTSGNGDPRYTNYRNNGHVEQKSAL YMRENNISNATVYHNNTNGTCGYCNTMTATFLPEGAT LTVVPPENAVANNSRAIDYVKTYTGTSNDPKISPRYKG N	152
	N2313-N	DPSGYDSQYPCKEEMSAGAGESGRKTISLPEYDGTTTH GVLVLDDGTQIGFTSGN	153
	N2313-C	GDPRYTNYRNNGHVEQKSALYMRENNISNATVYHNNT NGTCGYCNTMTATFLPEGATLTVVPPENAVANNSRAID YVKTYTGTSNDPKISPRYKGN	154
	R2385-N	DPSGYDSQYPCKEEMSAGAGESGRKTISLPEYDGTTTH GVLVLDDGTQIGFTSGNGDPRYTNYRNNGHVEQKSAL YMRENNISNATVYHNNTNGTCGYCNTMTATFLPEGAT LTVVPPENAVANNSR	155
	R2385-C	AIDYVKTYTGTSNDPKISPRYKGN	156
TDD7	toxic domain	MGDRLPAFVDGGDTLGIFSRGGIERDLASGVAGPASSL PKGTPGFNGLVKSHVEGHAAALMRQNGIPNAELYINR VPCGSGNGCAAMLPHMLPEGATLRVYGPNGYDRTFTG LPD	157
	G33-N	MGDRLPAFVDGGDTLGIFSRGGIERDLASGVAG	158
	G33-C	PASSLPKGTPGFNGLVKSHVEGHAAALMRQNGIPNAEL YINRVPCGSGNGCAAMLPHMLPEGATLRVYGPNGYDR TFTGLPD	159
	G102-N	MGDRLPAFVDGGDTLGIFSRGGIERDLASGVAGPASSL PKGTPGFNGLVKSHVEGHAAALMRQNGIPNAELYINR VPCGSGNGCAAMLPHMLPEGATLRVYG	160
	G102-C	PNGYDRTFTGLPD	161
TDD8	toxic domain	GGSAVVGAGVVATGAKAVTTGKSLSESQATLSVAQRL LATIGEEGKTAGVLELDGELIPLVSGKSSLPNYAASGHV EGQAALIMRDRGATSGRLLIDNPSGICGYCKSQVATLLP ENATLQVGTPLGTVTPSSRWSASRTFTGNDRDPKPWPR	162
	G2108-N	GGSAVVGAGVVATGAKAVTTGKSLSESQATLSVAQRL LATIGEEGKTAGVLELDGELIPLVSG	163
	G2108-C	KSSLPNYAASGHVEGQAALIMRDRGATSGRLLIDNPSGI CGYCKSQVATLLPENATLQVGTPLGTVTPSSRWSASRT FTGNDRDPKPWPR	164
	T2175-N	GGSAVVGAGVVATGAKAVTTGKSLSESQATLSVAQRL LATIGEEGKTAGVLELDGELIPLVSGKSSLPNYAASGHV EGQAALIMRDRGATSGRLLIDNPSGICGYCKSQVATLLP ENATLQVGTPLGTVT	165
	T2175-C	PSSRWSASRTFTGNDRDPKPWPR	166
TDD9	toxic domain	DIILATLPIGKVGKLRFAPKVESAESMLRSLSQEGKTAG	167

		VLDINGELIPLVSGTSSLKNYAASGHVEGQAALIMRER GVASARLIIDNPSGICGYCRSQVPTLLPAGATLEVTTPR GTVPPTARWSNGKTFVGNENDPKPWPR	
	G2112-N	DIILATLPIGKVGKLRFAPKVESAESMLRSLSQEGKTAG VLDINGELIPLVSG	168
	G2112-C	TSSLKNYAASGHVEGQAALIMRERGVASARLIIDNPSGI CGYCRSQVPTLLPAGATLEVTTPRGTVPPTARWSNGKT FVGNENDPKPWPR	169
	P2179-N	DIILATLPIGKVGKLRFAPKVESAESMLRSLSQEGKTAG VLDINGELIPLVSGTSSLKNYAASGHVEGQAALIMRER GVASARLIIDNPSGICGYCRSQVPTLLPAGATLEVTTPR GTVP	170
	P2179-C	PTARWSNGKTFVGNENDPKPWPR	171
TDD10	toxic domain	PPVASGGLATEVPAYAGSRTAGTLVTPDGAEFPLISGW HPPAASMPQGTPGMNIVTKSHVEAHAAAIMRNQGLSE ATLWINRAPCGGKPGCAAMLPRMVPSGSTLTINVVPNG SAGSIADTLIIRGIG	172
	G1667-N	PPVASGGLATEVPAYAGSRTAGTLVTPDGAEFPLISG	173
	G1667-C	WHPPAASMPQGTPGMNIVTKSHVEAHAAAIMRNQGLS EATLWINRAPCGGKPGCAAMLPRMVPSGSTLTINVVPN GSAGSIADTLIIRGIG	174
	G1746-N	PPVASGGLATEVPAYAGSRTAGTLVTPDGAEFPLISGW HPPAASMPQGTPGMNIVTKSHVEAHAAAIMRNQGLSE ATLWINRAPCGGKPGCAAMLPRMVPSGSTLTINVVPNG SAG	175
	G1746-C	SIADTLIIRGIG	176
TDD11	toxic domain	EIRAKYPTPEEAQLPPYDGDTTYALMYYTDEHGKSHV VELSSGGADDEHSNYAAAGHTEGQAAVIMRQRKITSA VVVHNNTDGTCPFCVAHLPTLLPSGAELRVVPPRSAKA KKPGWIDVSKTFEGNARKPLDNKNKKST	177
	G1430-N	EIRAKYPTPEEAQLPPYDGDTTYALMYYTDEHGKSHV VELSSGG	178
	G1430-C	ADDEHSNYAAAGHTEGQAAVIMRQRKITSAVVVHNNT DGTCPFCVAHLPTLLPSGAELRVVPPRSAKAKKPGWID VSKTFEGNARKPLDNKNKKST	179
	A1498-N	EIRAKYPTPEEAQLPPYDGDTTYALMYYTDEHGKSHV VELSSGGADDEHSNYAAAGHTEGQAAVIMRQRKITSA VVVHNNTDGTCPFCVAHLPTLLPSGAELRVVPPRSAKA	180
	A1498-C	KKPGWIDVSKTFEGNARKPLDNKNKKST	181
	G1502-N	EIRAKYPTPEEAQLPPYDGDTTYALMYYTDEHGKSHV VELSSGGADDEHSNYAAAGHTEGQAAVIMRQRKITSA VVVHNNTDGTCPFCVAHLPTLLPSGAELRVVPPRSAKA KKPG	182
	G1502-C	WIDVSKTFEGNARKPLDNKNKKST	183
TDD12	toxic domain	AALLREAYPSMEGATLPPFDGKTTIGLMFYTDASGQYQ VKKLFSGEKVLSNYDATGHVEGKAALIMRNEKITEAV VMHNHPSGTCNYCDKQVETLLPKNATLRVIPPENAKAP TSYWNDQPTTYRGDGKDPKAPSKK	184
	G1421-N	AALLREAYPSMEGATLPPFDGKTTIGLMFYTDASGQYQ VKKLFSG	185
	G1421-C	EKVLSNYDATGHVEGKAALIMRNEKITEAVVMHNHPS GTCNYCDKQVETLLPKNATLRVIPPENAKAPTSYWND QPTTYRGDGKDPKAPSKK	186
	A1488-N	AALLREAYPSMEGATLPPFDGKTTIGLMFYTDASGQYQ	187

		VKKLFSGEKVLSNYDATGHVEGKAALIMRNEKITEAV VMHNHPSGTCNYCDKQVETLLPKNATLRVIPPENAKA	
	A1488-C	PTSYWNDQPTTYRGDGKDPKAPSKK	188
TDD13	toxic domain	ALLREQFPSMDAVTLPPFDGKTTIGYMFYTDANGQYH VRKLYSGGKVLSNYDSSGHVEGMAALIMRKGRITEAV VMHNHPSGTCHYCNGQVETLLPKNAKLKVIPPANAKA PTKYWYDQPVDYLGNSNDPKPPS	189
	G1411-N	ALLREQFPSMDAVTLPPFDGKTTIGYMFYTDANGQYH VRKLYSGG	190
	G1411-C	KVLSNYDSSGHVEGMAALIMRKGRITEAVVMHNHPSG TCHYCNGQVETLLPKNAKLKVIPPANAKAPTKYWYDQ PVDYLGNSNDPKPPS	191
	A1477-N	ALLREQFPSMDAVTLPPFDGKTTIGYMFYTDANGQYH VRKLYSGGKVLSNYDSSGHVEGMAALIMRKGRITEAV VMHNHPSGTCHYCNGQVETLLPKNAKLKVIPPANAKA	192
	A1477-C	PTKYWYDQPVDYLGNSNDPKPPS	193
TDD14	toxic domain	GSSGKNVRMPRDYASELPEYDGKTTHGVLVTNEGKVI QLRSGGKEEPYTGYKAVSASHVEGKAAIWIRENGSSGG TVYHNNTTGTCGYCNSQVKALLPEGVELKIVPPTNAVA KNAQARAVPTINVGNGTQPGRKQK	194
	G43-N	GSSGKNVRMPRDYASELPEYDGKTTHGVLVTNEGKVI QLRSGG	195
	G43-C	KEEPYTGYKAVSASHVEGKAAIWIRENGSSGGTVYHN NTTGTCGYCNSQVKALLPEGVELKIVPPTNAVAKNAQ ARAVPTINVGNGTQPGRKQK	196
	A118-N	GSSGKNVRMPRDYASELPEYDGKTTHGVLVTNEGKVI QLRSGGKEEPYTGYKAVSASHVEGKAAIWIRENGSSGG TVYHNNTTGTCGYCNSQVKALLPEGVELKIVPPTNAVA KNAQA	197
	A118-C	RAVPTINVGNGTQPGRKQK	198
TDD15	toxic domain	GSSGKNVRLPRDYASELPEYDGKTTYGVLVTNEGKVIQ LRSGGKEVPYSGYKAVSASHVEGKAAIWIRENASSGGT VYHNNTTGTCGYCNSQVKALLPEGVELKIVPPANAVA RNSQAKAIPTINVGNATQPGRKP	199
	G315-N	GSSGKNVRLPRDYASELPEYDGKTTYGVLVTNEGKVIQ LRSGG	200
	G315-C	KEVPYSGYKAVSASHVEGKAAIWIRENASSGGTVYHN NTTGTCGYCNSQVKALLPEGVELKIVPPANAVARNSQA KAIPTINVGNATQPGRKP	201
	A390-N	GSSGKNVRLPRDYASELPEYDGKTTYGVLVTNEGKVIQ LRSGGKEVPYSGYKAVSASHVEGKAAIWIRENASSGGT VYHNNTTGTCGYCNSQVKALLPEGVELKIVPPANAVA RNSQA	202
	A390-C	KAIPTINVGNATQPGRKP	203
TDD16	toxic domain	PDPPPPPTPMGNTLPGWDGGKTQGWFVYPDGTERHLIS GYDGPSKFTQGIPGMNGNIKSHVEAHAAALMRQYELS KATLYINRVPCPGVRGCDALLARMLPEGVQLEIIGPNGF KKTYTGLPDPKLKPKGCS	204
	G1264-N	PDPPPPPTPMGNTLPGWDGGKTQGWFVYPDGTERHLIS GYDG	205
	G1264-C	PSKFTQGIPGMNGNIKSHVEAHAAALMRQYELSKATLY INRVPCPGVRGCDALLARMLPEGVQLEIIGPNGFKKTYT GLPDPKLKPKGCS	206
	G1342-C	PDPPPPPTPMGNTLPGWDGGKTQGWFVYPDGTERHLIS	207

		GYDGPSKFTQGIPGMNGNIKSHVEAHAAALMRQYELS KATLYINRVPCPGVRGCDALLARMLPEGVQLEIIGPNGF KKTYTG	
	G1342-N	LPDPKLKPKGCS	208
TDD17	toxic domain	GAATVFGAGRGLGALEEATTAAGIARGAPSLPVYTGG KTTGVLRTATGDMPLVSGYKGPSASMPRGTPGMNGRI KSHVEAHAAAVMRERGIKDATLHINQVPCSSATGCGA MLPRMLPEGAQLRVLGPDGYDQVFIGLPD	209
	G2087-N	GAATVFGAGRGLGALEEATTAAGIARGAPSLPVYTGG KTTGVLRTATGDMPLVSGYKG	210
	G2087-C	PSASMPRGTPGMNGRIKSHVEAHAAAVMRERGIKDAT LHINQVPCSSATGCGAMLPRMLPEGAQLRVLGPDGYD QVFIGLPD	211
	G2156-N	GAATVFGAGRGLGALEEATTAAGIARGAPSLPVYTGG KTTGVLRTATGDMPLVSGYKGPSASMPRGTPGMNGRI KSHVEAHAAAVMRERGIKDATLHINQVPCSSATGCGA MLPRMLPEGAQLRVLG	212
	G2156-C	PDGYDQVFIGLPD	213
TDD18	toxic domain	TNIIDNRPKLPDYDGKTTHGILVTPNSEHIPFSSGNPNPN YKNYIPASHVEGKSAIYMRENGITSGTIYYNNTDGTCPY CDKMLSTLLEEGSVLEVIPPINAKAPKPSWVDKPKTYIG NNKVPKPNK	214
	G181-N	TNIIDNRPKLPDYDGKTTHGILVTPNSEHIPFSSG	215
	G181-C	NPNPNYKNYIPASHVEGKSAIYMRENGITSGTIYYNNTD GTCPYCDKMLSTLLEEGSVLEVIPPINAKAPKPSWVDKP KTYIGNNKVPKPNK	216
	A250-N	TNIIDNRPKLPDYDGKTTHGILVTPNSEHIPFSSGNPNPN YKNYIPASHVEGKSAIYMRENGITSGTIYYNNTDGTCPY CDKMLSTLLEEGSVLEVIPPINAKA	217
	A250-C	PKPSWVDKPKTYIGNNKVPKPNK	218
TDD19	toxic domain	AGCPGDALPPYGTKGSKTTGILDTGNESILLESGENGPG MMVPRDTPGMSGAMPNRAHVEGHTAAIMRNENIRLA DLYINRMPCSGAYGCMVNLPHMLPEGSILRIHVRAKLS DPWTTLPPFVGISDTLWPPSGLNPKIVLP	219
	G234-N	AGCPGDALPPYGTKGSKTTGILDTGNESILLESGENG	220
	G234-C	PGMMVPRDTPGMSGAMPNRAHVEGHTAAIMRNENIRL ADLYINRMPCSGAYGCMVNLPHMLPEGSILRIHVRAKL SDPWTTLPPFVGISDTLWPPSGLNPKIVLP	221
	G321-N	ISDTLWPPSGLNPKIVLP	222
	G321-C	AGCPGDALPPYGTKGSKTTGILDTGNESILLESGENGPG MMVPRDTPGMSGAMPNRAHVEGHTAAIMRNENIRLA DLYINRMPCSGAYGCMVNLPHMLPEGSILRIHVRAKLS DPWTTLPPFVG	223

SEQ: SEQ ID NO:

TDD Base Editing Activity at the CCR5 *Locus*

[00148] **FIG. 12** shows the base editing frequency of TDD1-TDD6 (select splits) at C9, C10, C14, C16, C18, C20, and C24 of target sequence *CCR5*, with two different pairs of ZFP DNA binding domains (see **FIG. 10**). Two orientations of each split enzyme were tested (i.e., with the N- and C-terminal halves linked to different members of the ZFP pair for each

orientation). In experiments where the base editing system included a nickase, a ZFP-FokI nickase or a CRISPR/Cas9 nickase was used.

[00149] **FIG. 13** shows a comparison of the highest frequency of editing for each deaminase for any C in the base editing window (based on data shown in **FIG. 12** as well as additional replicates). At least three of the TDDs (TDD3, TDD4, and TDD6) demonstrated detectable base editing activity (>0.25% base editing), with TDD4 showing higher maximum activity than DddA.

[00150] **FIG. 14** provides a more detailed analysis of the TDD base editing activity (based on data shown in **FIG. 12** as well as additional replicates), showing the highest frequency of editing for any C in the base editing window for the two binding orientations of each TDD to the two different ZFP pairs, with or without nickase activity. Base editing for certain TDDs appeared to be sensitive to the ZFP pair (e.g., TDD4) or the binding orientation (e.g., TDD3). TDD6 seemed to have detectable activity (>0.25% base editing) for every condition under which it was tested, albeit with a binding orientation dependency at least in the context of ZFP#4 and ZFP#5. For each TDD, in some cases, nicking appeared to improve base editing activity (see also **FIG. 12**).

TDD Base Editing Activity at the CIITA *Locus*

[00151] Select TDD split enzymes were tested for base editing at the nucleotides labeled G2, G5, C6, C8, G10, G11, G14, C15 and C16 in target sequence *CIITA* with the ZFP binding domains shown ("CIITA_site_2_right_1," "CIITA_site_2_right_5," and "CIITA_site_2_left_6") (**FIG. 15**). **FIG. 16** shows a comparison of the highest frequency of editing for each fusion protein pair for any C in the base editing window. TDD3, TDD4, and TDD6, which were active at the *CCR5* locus, also demonstrated detectable base editing activity (>0.25% base editing) at the *CIITA* locus. Eight additional TDDs (TDD8, TDD9, TDD10, TDD12, TDD14, TDD15, TDD18, and TDD19) demonstrated detectable editing as well. Base editing activity appeared to be sensitive to the TDD split position, and in some cases to the variant of the toxic domain used (e.g., TDD4). TDD4 appeared to have significant activity in every condition under which it was tested. Some TDDs also provide an increased targeting density (**FIG. 17**) with stronger activity at TC and AC sites (compared to DddA; *see, e,g.*, TDD6) as well as activity at GC and CC sites (e.g., TDD6).

Effect of Different Linkers on TDD Base Editing Activity at the CIITA *Locus*

[00152] To assess whether base editing activity is affected by different linkers between the deaminase and ZFP domains, the editing frequency of TDD6 at the *CIITA* locus was assessed with linkers L26, L21, L18, L13, L11, L9, L6, and L4. As shown in **FIG. 18**, different linker lengths were able to alter the base editing profile within the base editing window. For example, shortening the linker connecting the left ZFP to either the N- or C-terminal TDD split appeared to narrow the activity window. Such alterations may increase base editor precision and specificity. In some cases, the effects of linker length appeared sensitive to the binding orientation of the TDD splits to the ZFP pair or to the TDD (e.g., L4 performance with TDD14).

Example 7: Targeting inhibitor TDDI to TDD

[00153] TDD enzymes may be inactivated by TDDIs. For example, the natural DddA enzyme can be inactivated by the DddI inhibitor. A ZFP or TALE linked TDDI can be targeted to a potential TDD-derived cytosine base editor site, preventing that site from being edited (**FIG. 19**). The TDDI inhibitor may be linked to the ZFP using a dimerization domain potentiated by a small molecule, thus putting the editing activity under the control of the small molecule.

[00154] By designing the targeted TDDI construct to be allele specific, editing can selectively be targeted to certain alleles, e.g., to knock out a detrimental mutant by editing in a stop codon only if the mutation is present. For example, JAK2 V617F can be knocked out by editing in a stop codon only if the V617F mutation is present.

[00155] This TDDI approach may also be used to reduce editing at off-target sites, particularly where it cannot be eliminated by other means.

[00156] It is also contemplated that other cytidine deaminases and their inhibitors can be used in place of a TDD and TDDI.

What is claimed is:

1. A system for changing a cytosine to a thymine in the genome of a cell, comprising a first fusion protein and a second fusion protein, or first and second expression constructs for expressing the first and second fusion proteins, respectively, wherein

a) the first fusion protein comprises:

i) a first zinc finger protein (ZFP) domain that binds to a first sequence in a target genomic region in the cell, and

ii) a first portion of a cytidine deaminase polypeptide, wherein the cytidine deaminase is a toxin-derived deaminase (TDD) comprising an amino acid sequence at least 90% identical to SEQ ID NO: 49, 81, 92, 95, 98, 101, 104, 107, 134, 143, 152, 157, 162, 167, 172, 177, 184, 189, 194, 199, 204, 209, 214, or 219;

b) the second fusion protein comprises:

i) a second ZFP domain that binds to a second sequence in the target genomic region, and

ii) a second portion of the cytidine deaminase polypeptide;

c) the first and second portions lack cytidine deaminase activity on their own; and

d) binding of the first fusion protein and the second fusion protein to the target genomic region results in dimerization of the first and second portions, wherein the dimerized portions form an active cytidine deaminase capable of changing a cytosine to a thymine in the target genomic region,

optionally wherein the cell is a eukaryotic cell,

optionally wherein the eukaryotic cell is a mammalian cell or a plant cell,

further optionally wherein the mammalian cell is a human cell.

2. The system of claim 1, wherein the target genomic region is specific to a particular allele of a gene in the cell.

3. The system of claim 1 or 2, wherein the cytosine is between the proximal ends of the first sequence and the second sequence in the target genomic region, optionally wherein the proximal ends are no more than 100 bps apart.

4. The system of any one of claims 1-3, comprising more than one pair of the first and second fusion proteins, wherein each pair of the fusion proteins binds to a different target genomic region.

5. The system of claim 4, wherein the first and second cytidine deaminase portions of one pair of fusion proteins are different from the first and second portions of another pair of fusion proteins.

6. The system of any one of claims 1-5, further comprising a nickase that creates a single-stranded DNA break on the unedited or edited strand, wherein the DNA break is no more than about 500 bps, optionally no more than 200 bps, optionally about 10-50 bps, from the cytosine to be edited.

7. The system of claim 6, wherein the nickase is a ZFP-based nickase, a TALE-based nickase, or a CRISPR-based nickase.

8. The system of claim 7, wherein the nickase is a ZFP-based nickase formed by dimerization of a first nickase domain and a second nickase domain fused respectively to two ZFP domains that bind to the target genomic region, wherein the first and second nickase domains are inactive on their own.

9. The system of claim 8, wherein

one of the nickase domains is fused to the first or second fusion protein, and

the other nickase domain is fused to a third ZFP domain that binds to a third sequence in the target genomic region.

10. The system of claim 8, wherein the two nickase domains are fused respectively to

i) a third ZFP domain that binds a third sequence in the target genomic region and

ii) a fourth ZFP domain that binds a fourth sequence in the target genomic region.

11. The system of any one of claims 8-10, wherein the first and second nickase domains are derived from FokI.

12. The system of any one of claims 1-7, further comprising a third fusion protein or a third expression construct for expressing the third fusion protein in the cell, wherein

e) the third fusion protein comprises

i) a ZFP domain that binds to a third sequence in the target genomic region, and

ii) an inhibitory domain for the cytidine deaminase; and

f) binding of the third fusion protein to the target genomic region results in the inhibitory domain binding to, and thereby inhibition of the cytidine deaminase activity of, the dimerized cytidine deaminase portions.

13. The system of any one of claims 1-7, further comprising a third fusion protein or a third expression construct for expressing the third fusion protein in the cell, and a fourth fusion protein or a fourth expression construct for expressing the fourth fusion protein in the cell, wherein

e) the third fusion protein comprises

i) a ZFP domain that binds to a third sequence in the target genomic region, and

ii) a first dimerization domain; and

f) the fourth fusion protein comprises

i) an inhibitory domain for the cytidine deaminase, and

ii) a second dimerization domain capable of partnering with the first dimerization domain in the presence of a dimerization-inducing agent; and

g) binding of the third fusion protein to the target genomic region, and dimerization of the first and second dimerization domains, result in the inhibitory domain binding to, and thereby inhibition of the cytidine deaminase activity of, the dimerized cytidine deaminase portions.

14. The system of any one of claims 1-7, further comprising a third fusion protein or a third expression construct for expressing the third fusion protein in the cell, and a fourth fusion protein or a fourth expression construct for expressing the fourth fusion protein in the cell, wherein

e) the third fusion protein comprises

i) a ZFP domain that binds to a third sequence in the target genomic region, and

ii) a first dimerization domain; and

f) the fourth fusion protein comprises

i) an inhibitory domain for the cytidine deaminase, and

ii) a second dimerization domain capable of partnering with the first dimerization domain in the absence of a dimerization-inhibiting agent; and

g) binding of the third fusion protein to the target genomic region, and dimerization of the first and second dimerization domains, result in the inhibitory domain binding to, and thereby inhibition of the cytidine deaminase activity of, the dimerized cytidine deaminase portions.

15. The system of any one of the preceding claims, wherein the ZFP domains independently have 2, 3, 4, 5, 6, 7, or 8 zinc fingers.

16. The system of any one of the preceding claims, wherein the expression constructs are on the same or separate viral vectors.

17. The system of claim 16, wherein the viral vectors are adeno-associated viral (AAV) vectors, adenoviral vectors, or lentiviral vectors.

18. The system of any one of claims 1-17, wherein the TDD comprises the amino acid sequence of SEQ ID NO: 72.

19. The system of any one of claims 1-17, wherein the TDD comprises the toxic domain of a TDD comprising the amino acid sequence of SEQ ID NO: 72.

20. The system of any one of claims 1-17, wherein the cytidine deaminase is a TDD that comprises an amino acid sequence at least 95% identical to the amino acid sequence of SEQ ID NO: 49 or 81.

21. The system of any one of claims 1-17, wherein the TDD comprises the amino acid sequence of SEQ ID NO: 49 or 81.

22. The system of any one of claims 1-17, wherein

the first and second cytidine deaminase portions comprise:

amino acids 1264-1333 and 1334-1427, respectively;
amino acids 1264-1397 and 1398-1427, respectively;
amino acids 1264-1404 and 1405-1427, respectively;
amino acids 1264-1407 and 1408-1427, respectively;
amino acids 1290-1333 and 1334-1427, respectively;
amino acids 1290-1397 and 1398-1427, respectively;
amino acids 1290-1404 and 1405-1427, respectively; or
amino acids 1290-1407 and 1408-1427, respectively;
of SEQ ID NO: 72; or
vice versa.

23. The system of any one of claims 1-17, wherein:
the first and second cytidine deaminase portions respectively comprise
SEQ ID NOs: 82 and 83,
SEQ ID NOs: 84 and 85,
SEQ ID NOs: 18 and 19,
SEQ ID NOs: 51 and 52, or
SEQ ID NOs: 53 and 54; or
vice versa.

24. The system of any one of claims 18-23, wherein the TDD has a mutation at one or more residues selected from Y1307, T1311, S1331, V1346, H1366, N1367, N1368, P1369, E1370, G1371, T1372, F1375, V1392, P1394, P1395, I1399, P1400, V1401, K1402, A1405, and T1406, wherein the residues are numbered with respect to SEQ ID NO: 72.

25. The system of any one of claims 1-17, wherein the cytidine deaminase is a TDD that comprises the amino acid sequence of any one of SEQ ID NOs: 86-91 and 117-129.

26. The system of any one of claims 1-17, wherein the cytidine deaminase comprises the toxic domain of a TDD comprising the amino acid sequence of any one of SEQ ID NOs: 86-91 and 117-129.

27. The system of any one of claims 1-17, wherein the TDD comprises an amino acid sequence at least 95% identical to the amino acid sequence of SEQ ID NO: 92, 95, 98, 101, 104, 107, 134, 143, 152, 157, 162, 167, 172, 177, 184, 189, 194, 199, 204, 209, 214, or 219.

28. The system of any one of claims 1-17, wherein the cytidine deaminase is a TDD that comprises the amino acid sequence of SEQ ID NO: 92, 95, 98, 101, 104, 107, 134, 143, 152, 157, 162, 167, 172, 177, 184, 189, 194, 199, 204, 209, 214, or 219.

29. The system of any one of claims 1-17, wherein

the first and second cytidine deaminase portions respectively comprise SEQ ID NOs: 93 and 94, SEQ ID NOs: 96 and 97, SEQ ID NOs: 99 and 100, SEQ ID NOs: 102 and 103, SEQ ID NOs: 105 and 106, SEQ ID NOs: 108 and 109, SEQ ID NOs: 130 and 131, SEQ ID NOs: 132 and 133, SEQ ID NOs: 135 and 136, SEQ ID NOs: 137 and 138, SEQ ID NOs: 139 and 140, SEQ ID NOs: 141 and 142, SEQ ID NOs: 144 and 145, SEQ ID NOs: 146 and 147, SEQ ID NOs: 148 and 149, SEQ ID NOs: 150 and 151, SEQ ID NOs: 153 and 154, SEQ ID NOs: 155 and 156, SEQ ID NOs: 158 and 159, SEQ ID NOs: 160 and 161, SEQ ID NOs: 163 and 164, SEQ ID NOs: 165 and 166, SEQ ID NOs: 168 and 169, SEQ ID NOs: 170 and 171, SEQ ID NOs: 173 and 174, SEQ ID NOs: 175 and 176, SEQ ID NOs: 178 and 179, SEQ ID NOs: 180 and 181, SEQ ID NOs: 182 and 183, SEQ ID NOs: 185 and 186, SEQ ID NOs: 187 and 188, SEQ ID NOs: 190 and 191, SEQ ID NOs: 192 and 193, SEQ ID NOs: 195 and 196, SEQ ID NOs: 197 and 198, SEQ ID NOs: 200 and 201, SEQ ID NOs: 202 and 203, SEQ ID NOs: 205 and 206, SEQ ID NOs: 207 and 208, SEQ ID NOs: 210 and 211, SEQ ID NOs: 212 and 213, SEQ ID NOs: 215 and 216, SEQ ID NOs: 217 and 218, SEQ ID NOs: 220 and 221, or SEQ ID NOs: 222 and 223; or

vice versa.

30. A fusion protein comprising i) a zinc finger protein (ZFP) domain that binds to a gene, and ii) a fragment of a cytidine deaminase polypeptide, wherein the cytidine deaminase is a toxin-derived deaminase (TDD) comprising an amino acid sequence at least 90% identical to SEQ ID NO: 49, 81, 92, 95, 98, 101, 104, 107, 134, 143, 152, 157, 162, 167, 172, 177, 184, 189, 194, 199, 204, 209, 214, or 219, optionally wherein the ZFP domain and the cytidine deaminase fragment are linked by a peptide linker, optionally wherein the gene is a eukaryotic gene, optionally wherein the eukaryotic gene is a human gene.

31. A fusion protein comprising i) a zinc finger protein (ZFP) domain that binds to a gene, and ii) a cytidine deaminase inhibitory domain, wherein the cytidine deaminase is a toxin-derived deaminase (TDD) comprising an amino acid sequence at least 90% identical to SEQ ID NO: 49, 81, 92, 95, 98, 101, 104, 107, 134, 143, 152, 157, 162, 167, 172, 177, 184, 189, 194, 199, 204, 209, 214, or 219, optionally wherein the ZFP domain and the inhibitory domain are linked by a peptide linker, optionally wherein the gene is a eukaryotic gene, optionally wherein the eukaryotic gene is a human gene.

32. The fusion protein of claim 30 or 31, wherein the TDD comprises the amino acid sequence of SEQ ID NO: 49, 81, 92, 95, 98, 101, 104, 107, 134, 143, 152, 157, 162, 167, 172, 177, 184, 189, 194, 199, 204, 209, 214, or 219.

33. The fusion protein of any one of claims 30-32, wherein the linker comprises any one of SEQ ID NOs: 15-17 and 110-116.

34. A pair of fusion proteins comprising

a) a first fusion protein that comprises i) a zinc finger protein (ZFP) domain that binds to a gene, and ii) a first dimerization domain, and

b) a second fusion protein that comprises i) a cytidine deaminase inhibitory domain, wherein the cytidine deaminase is a toxin-derived deaminase (TDD) comprising an amino acid sequence at least 90% identical to SEQ ID NO: 49, 81, 92, 95, 98, 101, 104, 107, 134, 143, 152, 157, 162, 167, 172, 177, 184, 189, 194, 199, 204, 209, 214, or 219, and ii) a second dimerization domain,

wherein the first and second dimerization domains can dimerize in the presence of a dimerization-inducing agent,

optionally wherein the gene is a eukaryotic gene, optionally wherein the eukaryotic gene is a human gene.

35. A pair of fusion proteins comprising

a) a first fusion protein that comprises i) a zinc finger protein (ZFP) domain that binds to a gene, and ii) a first dimerization domain, and

b) a second fusion protein that comprises i) a cytidine deaminase inhibitory domain, wherein the cytidine deaminase is a toxin-derived deaminase (TDD) comprising an amino acid sequence at least 90% identical to SEQ ID NO: 49, 81, 92, 95, 98, 101, 104, 107, 134,

143, 152, 157, 162, 167, 172, 177, 184, 189, 194, 199, 204, 209, 214, or 219, and ii) a second dimerization domain,

wherein the first and second dimerization domains can dimerize in the absence of a dimerization-inhibiting agent,

optionally wherein the gene is a eukaryotic gene, optionally wherein the eukaryotic gene is a human gene.

36. The pair of fusion proteins of claim 34 or 35, wherein the TDD comprises the amino acid sequence of SEQ ID NO: 49, 81, 92, 95, 98, 101, 104, 107, 134, 143, 152, 157, 162, 167, 172, 177, 184, 189, 194, 199, 204, 209, 214, or 219.

37. One or more isolated nucleic acid molecules encoding the fusion protein(s) of any one of claims 30-36.

38. An expression construct comprising the nucleic acid molecule(s) of claim 37.

39. A viral vector comprising the expression construct of claim 38, optionally wherein the viral vector is an adeno-associated viral vector, an adenoviral vector, or a lentiviral vector.

40. A cell comprising the system of any one of claims 1-29, the fusion protein(s) of any one of claims 30-36, the isolated nucleic acid molecule(s) of claim 37, the expression construct of claim 38, or the viral vector of claim 39, optionally wherein the cell is a eukaryotic cell.

41. The cell of claim 40, wherein the cell is a mammalian cell, optionally a human cell, further optionally a human embryonic stem or a human induced pluripotent stem cell.

42. A method of changing a cytosine to a thymine in a target genomic region in a cell, comprising delivering the system of any one of claims 1-29 to the cell, optionally wherein the cell is a eukaryotic cell.

43. The method of claim 42, wherein the change of the cytosine to the thymine creates a stop codon in the target genomic region.

44. The method of claim 42 or 43, wherein the system targets more than one genomic region.

45. The method of any one of claims 42-44, comprising delivering the system of any one of claims 13 and 15-29 and the dimerization-inducing agent, wherein the agent induces dimerization of the first and second dimerization domains and thereby activates binding of the inhibitory domain to the dimerized cytidine deaminase portions.

46. The method of any one of claims 42-44, comprising delivering the system of any one of claims 14-29 and the dimerization-inhibiting agent, wherein the agent inhibits dimerization of the first and second dimerization domains and thereby prevents binding of the inhibitory domain to the dimerized cytidine deaminase portions.

47. The method of any one of claims 42-46, wherein the cell is a human cell *in vivo*.

48. The method of any one of claims 42-46, wherein the cell is a human cell *ex vivo*.

49. A genetically engineered cell, optionally a eukaryotic cell, optionally a human cell, obtained by the method of claim 48.

50. A method of treating a patient in need thereof, comprising delivering the genetically engineered cell of claim 49 to the patient, optionally wherein the cell and the patient are human.

51. The genetically engineered cell of claim 49, for use in treating a patient in need thereof.

52. Use of the genetically engineered cell of claim 49 for the manufacture of a medicament for treating a patient in need thereof.

53. The method, cell, or use of any one of claims 50-52, wherein the patient has cancer, an autoimmune disorder, an autosomal dominant disease, or a mitochondrial disorder.

54. The method, cell, or use of any one of claims 50-52, wherein the patient has sickle cell disease, hemophilia, cystic fibrosis, phenylketonuria, Tay-Sachs, prion disease, color blindness, a lysosomal storage disease, Friedreich's ataxia, or prostate cancer.

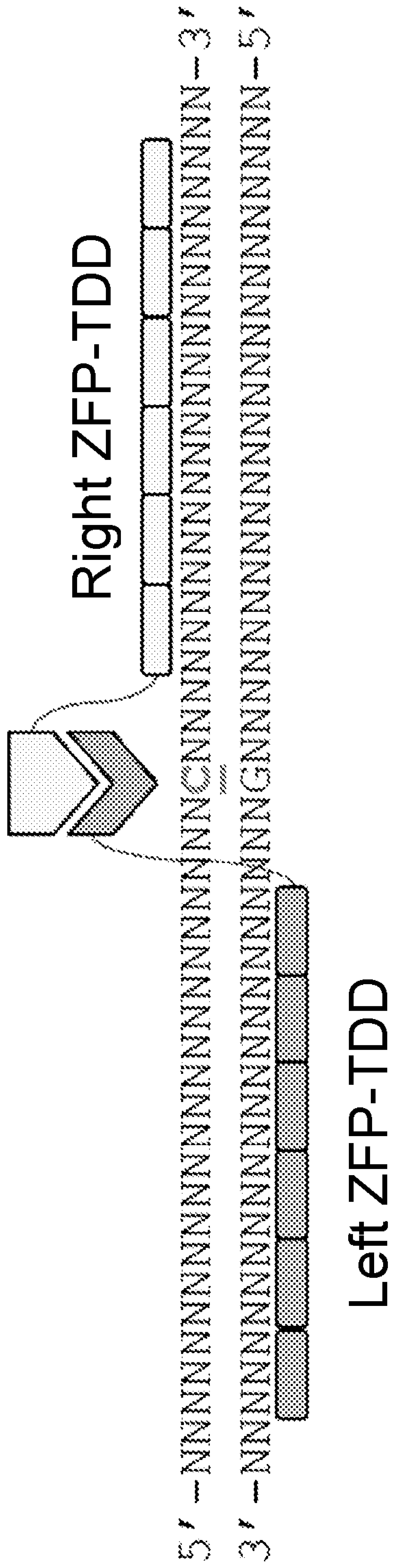


FIG. 1

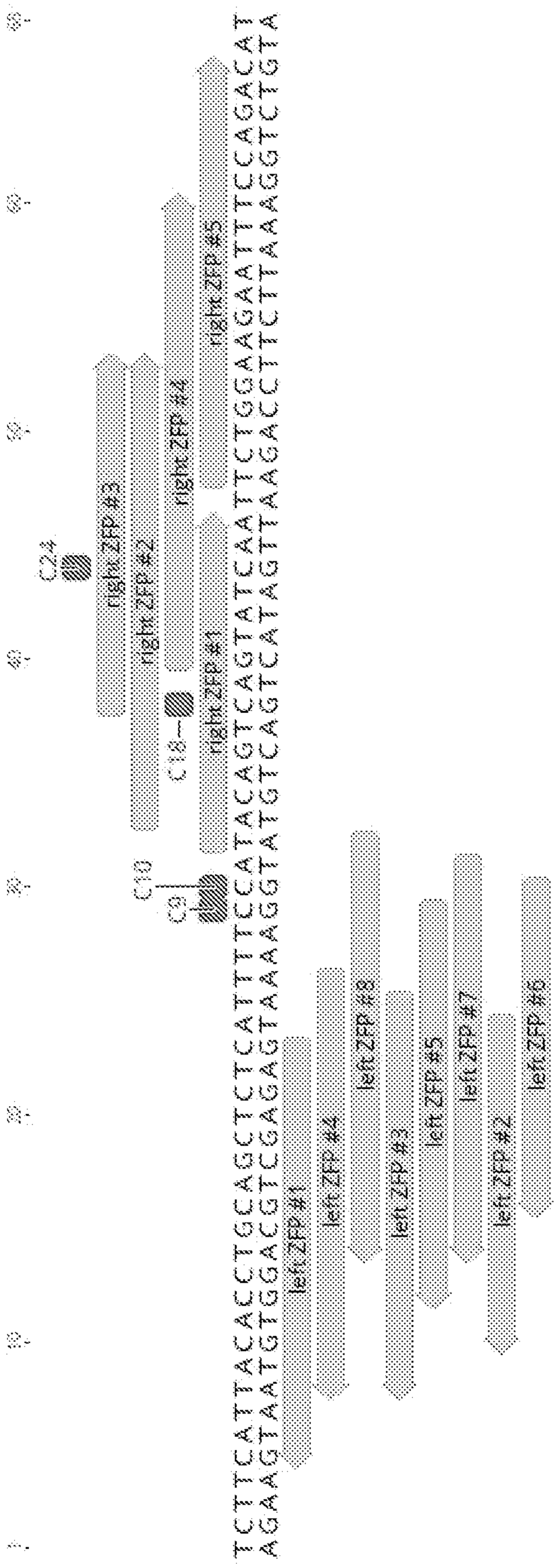

5'
60
50
C24
right ZFP #3
right ZFP #2
18
right ZFP #4
right ZFP #1
right ZFP #5
GTCAGTATCAATTCTTGGAAGAATTTC
CAGTCATAGTTAAGACCTTCTTAAAG
40
30
C10
C9
left ZFP #1
left ZFP #4
left ZFP #8
left ZFP #3
left ZFP #5
left ZFP #7
left ZFP #2
left ZFP #6
CCTGCAGCTCTCATTTTCCATACA
GGACGTCGAGAGTAAAAGGTATGT
20
10
TCTTCATTACA
AGAAGTAATGT
3'

FIG. 2A

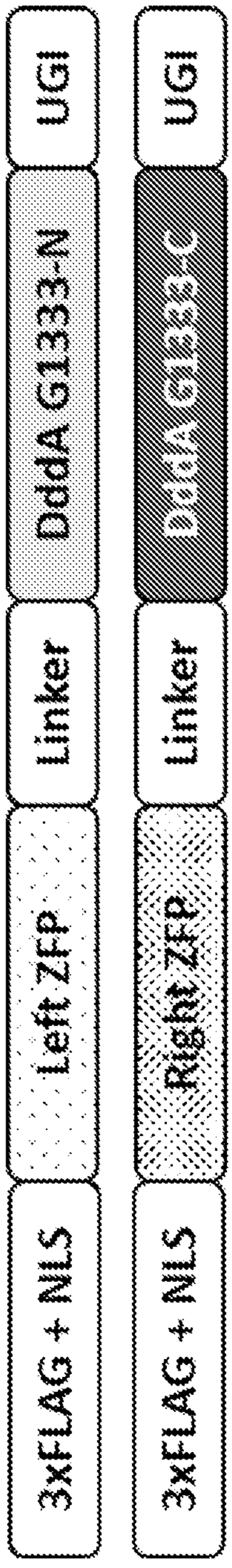

3xFLAG + NLS
Linker
Left ZFP
DddA G1333-N
UGI
3xFLAG + NLS
Linker
Right ZFP
DddA G1333-C
UGI

FIG. 2B

Left ZFP-DddA		Right ZFP-DddA			Base Editing % (C>T)			
ZFP	DddA	ZFP	DddA	Linker	C9	C10	C18	C24
Left ZFP #4	G1333-N	Right ZFP #4	G1333-C	L0	0.22	0.07	1.36	0.07
Left ZFP #4	G1333-N	Right ZFP #5	G1333-C	L0	0.05	0.04	0.38	1.02
Left ZFP #1	G1333-N	Right ZFP #1	G1333-C	L7A	2.98	1.24	0.03	0.03
Left ZFP #5	G1333-N	Right ZFP #5	G1333-C	L7A	0.04	0.03	1.05	0.32
Left ZFP #8	G1333-N	Right ZFP #4	G1333-C	L7A	0.04	0.03	3.32	0.04
Left ZFP #8	G1333-N	Right ZFP #5	G1333-C	L7A	0.03	0.04	3.02	1.12
Left ZFP #3	G1333-C	Right ZFP #1	G1333-N	L7A	1.11	0.94	0.04	0.03
Left ZFP #1	G1333-N	Right ZFP #1	G1333-C	L26	2.08	0.98	0.06	0.03
Left ZFP #3	G1333-N	Right ZFP #5	G1333-C	L26	0.06	0.04	0.21	1.70
Left ZFP #4	G1333-N	Right ZFP #5	G1333-C	L26	0.05	0.04	0.65	4.26
Left ZFP #5	G1333-N	Right ZFP #4	G1333-C	L26	0.03	0.03	1.03	0.05
Left ZFP #5	G1333-N	Right ZFP #5	G1333-C	L26	0.04	0.03	1.48	1.02
Left ZFP #8	G1333-N	Right ZFP #4	G1333-C	L26	0.03	0.04	2.12	0.03
Left ZFP #8	G1333-N	Right ZFP #5	G1333-C	L26	0.03	0.03	3.08	0.60
Left ZFP #3	G1333-C	Right ZFP #1	G1333-N	L26	1.02	0.87	0.04	0.04
Left ZFP #4	G1333-C	Right ZFP #1	G1333-N	L26	3.27	3.11	0.04	0.03
Left ZFP #4	G1333-C	Right ZFP #5	G1333-N	L26	0.13	0.07	2.04	0.21
Left ZFP #8	G1333-C	Right ZFP #5	G1333-N	L26	0.04	0.05	1.34	1.24

FIG. 3

		Base Editing %	
		C18	C24
Left_ZFP#8-G1397-C	Right_ZFP#5-G1397-N	3.78	2.05
Left_ZFP#8-G1404-C	Right_ZFP#5-G1404-N	22.98	0.07
Left_ZFP#8-G1407-C	Right_ZFP#5-G1407-N	17.03	0.04

FIG. 4

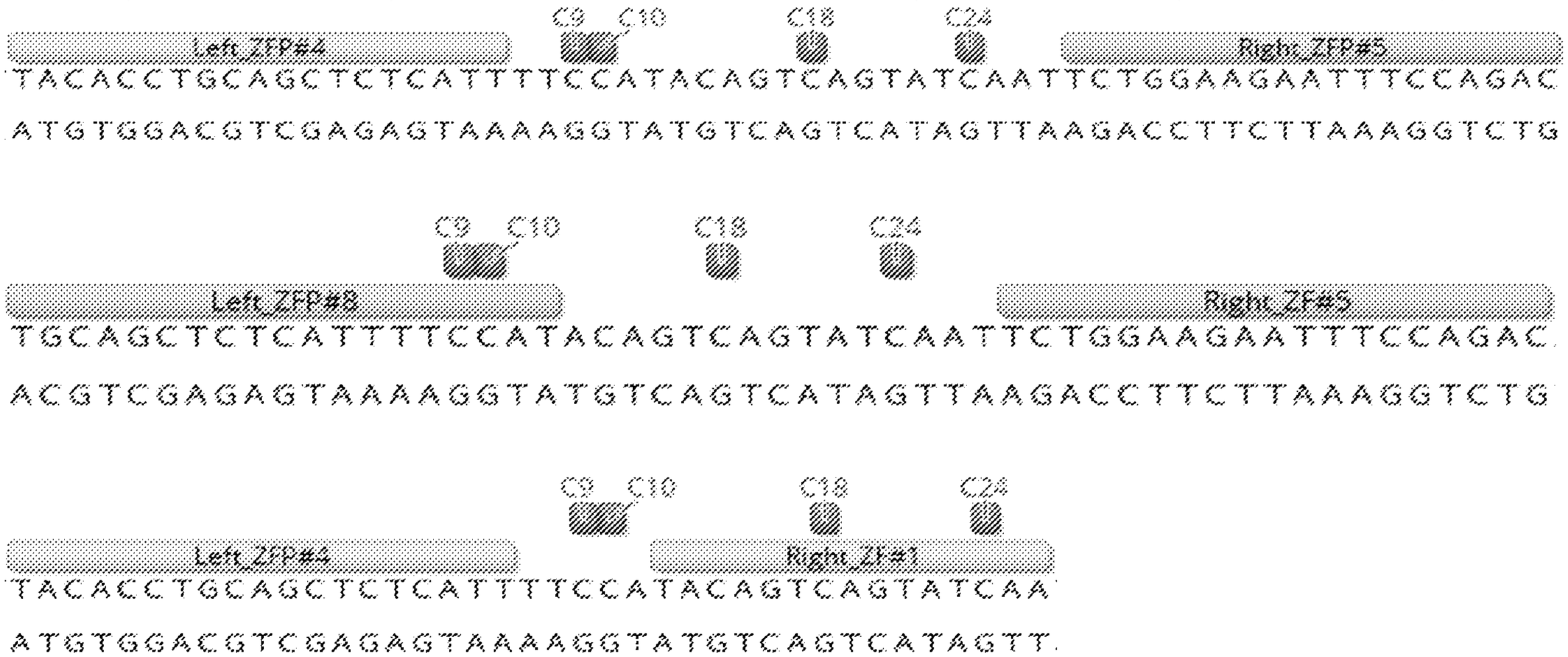

C9
C10
C18
C24
Left_ZFP#4
Right_ZFP#5
TACACCTGCAGCTCTCATTTTCCATACAGTCAGTATCAATTCTGGAAGAATTTCCAGAC
ATGTGGACGTCGAGAGTAAAAGGTATGTCAGTCATAGTTAAGACCTTCTTAAAGGTCTG
C9
C10
C18
C24
Left_ZFP#3
Right_ZF#3
TGCAGCTCTCATTTTCCATACAGTCAGTATCAATTCTGGAAGAATTTCCAGAC
ACGTCGAGAGTAAAAGGTATGTCAGTCATAGTTAAGACCTTCTTAAAGGTCTG
C9
C10
C18
C24
Left_ZFP#4
Right_ZF#1
TACACCTGCAGCTCTCATTTTCCATACAGTCAGTATCAA
ATGTGGACGTCGAGAGTAAAAGGTATGTCAGTCATAGTT

FIG. 5

	Left_ZFP#4-G1333-N : Right_ZFP#5-G1333-C				Left_ZFP#8-G1333-N : Right_ZFP#5-G1333-C				Left_ZFP#4-G1333-N : Right_ZFP#1-G1333-C			
	Base Editing %				Base Editing %				Base Editing %			
	C9	C10	C18	C24	C9	C10	C18	C24	C9	C10	C18	C24
WT	0.04	0.05	0.46	3.59	0.05	0.05	4.63	1.37	4.87	4.65	0.08	0.04
Y1307_Ala	0.02	0.05	0.06	0.14	0.07	0.04	0.71	0.17	0.05	0.07	0.05	0.03
Y1307_Arg	0.02	0.05	0.05	0.16	0.03	0.04	1.71	0.50	0.05	0.05	0.04	0.03
Y1307_Asn	0.04	0.06	0.04	0.04	0.04	0.05	0.07	0.05	0.12	0.08	0.04	0.03
Y1307_Asp	0.03	0.04	0.04	0.03	0.03	0.02	0.06	0.13	0.02	0.03	0.02	0.02
Y1307_Cys	0.03	0.05	0.05	0.15	0.04	0.04	0.29	0.15	0.13	0.12	0.02	0.02
Y1307_Gln	0.03	0.03	0.04	0.08	0.04	0.05	0.11	0.13	0.04	0.06	0.05	0.02
Y1307_Glu	0.04	0.04	0.03	0.05	0.02	0.04	0.05	0.09	0.04	0.05	0.04	0.02
Y1307_Gly	0.04	0.03	0.05	0.09	0.03	0.03	0.11	0.14	0.04	0.04	0.03	0.02
Y1307_His	0.04	0.05	0.08	0.70	0.03	0.05	0.69	0.21	0.40	0.27	0.04	0.02
Y1307_Ile	0.03	0.03	0.04	0.06	0.03	0.05	0.10	0.10	0.05	0.06	0.03	0.03
Y1307_Leu	0.03	0.05	0.04	0.11	0.04	0.03	0.11	0.11	0.03	0.06	0.05	0.03
Y1307_Lys	0.02	0.05	0.04	0.16	0.03	0.04	0.08	0.08	0.05	0.04	0.05	0.03
Y1307_Met	0.04	0.03	0.05	0.10	0.02	0.04	0.25	0.10	0.21	0.28	0.04	0.03
Y1307_Phe	1.01	0.97	9.18	3.20	0.04	0.04	2.22	0.36	4.07	4.21	0.06	0.04
Y1307_Pro	NA	NA	NA	NA	0.02	0.03	0.13	0.10	0.11	0.10	0.04	0.03
Y1307_Ser	NA	NA	NA	NA	0.03	0.06	0.09	0.07	0.06	0.07	0.04	0.03
Y1307_Trp	NA	NA	NA	NA	0.03	0.04	1.85	0.29	3.41	3.62	0.06	0.03
Y1307_Val	0.09	0.14	0.03	0.11	0.03	0.06	2.14	0.33	0.05	0.06	0.04	0.04
Y1307_Thr	0.02	0.04	0.04	0.17	0.03	0.04	0.17	0.11	0.07	0.05	0.03	0.02

FIG. 6A

	Left_ZFP#4-G1333-N : Right_ZFP#5-G1333-C				Left_ZFP#8-G1333-N : Right_ZFP#5-G1333-C				Left_ZFP#4-G1333-N : Right_ZFP#1-G1333-C			
	Base Editing %				Base Editing %				Base Editing %			
	C9	C10	C18	C24	C9	C10	C18	C24	C9	C10	C18	C24
WT	0.04	0.05	0.46	3.59	0.05	0.05	4.63	1.37	4.87	4.65	0.08	0.04
N1368_Ala	0.03	0.04	0.04	0.11	0.03	0.04	0.13	0.05	0.02	0.04	0.07	0.03
N1368_Arg	0.04	0.03	0.03	0.06	0.04	0.04	0.09	0.03	0.03	0.05	0.03	0.03
N1368_Asp	0.05	0.04	0.06	0.12	0.04	0.04	0.09	0.02	0.04	0.04	0.03	0.03
N1368_Cys	0.03	0.04	0.14	1.11	0.03	0.05	0.82	0.16	0.12	0.09	0.04	0.02
N1368_Gln	0.03	0.03	0.05	0.31	0.03	0.05	0.37	0.09	0.03	0.05	0.02	0.03
N1368_Glu	0.03	0.04	0.03	0.03	0.03	0.04	0.05	0.03	0.02	0.03	0.03	0.04
N1368_Gly	0.03	0.03	0.04	0.07	0.03	0.04	0.07	0.03	0.04	0.03	0.04	0.03
N1368_His	0.04	0.05	0.05	0.66	0.02	0.03	0.53	0.11	0.23	0.14	0.04	0.04
N1368_Ile	0.02	0.05	0.04	0.02	0.03	0.05	0.06	0.03	0.03	0.04	0.03	0.02
N1368_Leu	0.04	0.04	0.04	0.02	0.03	0.04	0.03	0.02	0.04	0.05	0.03	0.03
N1368_Lys	0.04	0.03	0.05	0.06	0.03	0.05	0.06	0.03	0.04	0.05	0.04	0.03
N1368_Met	0.03	0.04	0.06	0.21	0.03	0.03	0.26	0.07	0.06	0.04	0.03	0.02
N1368_Phe	0.03	0.04	0.04	0.09	0.04	0.04	0.08	0.04	0.04	0.03	0.01	0.12
N1368_Pro	0.03	0.03	0.02	0.02	0.05	0.05	0.03	0.01	NA	NA	NA	NA
N1368_Ser	0.03	0.04	0.05	0.11	0.03	0.04	0.1	0.03	0.03	0.04	0.03	0.02
N1368_Thr	0.02	0.05	0.03	0.04	0.04	0.04	0.04	0.02	0.02	0.05	0.05	0.03
N1368_Trp	0.02	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.02
N1368_Tyr	0.04	0.05	0.04	0.04	0.04	0.05	0.08	0.03	0.03	0.04	0.04	0.02
N1368_Val	0.03	0.04	0.05	0.03	0.03	0.05	0.04	0.03	0.03	0.05	0.03	0.04

FIG. 6B

	Left_ZFP#4-G1333-N : Right_ZFP#5-G1333-C				Left_ZFP#8-G1333-N : Right_ZFP#5-G1333-C				Left_ZFP#4-G1333-N : Right_ZFP#1-G1333-C			
	Base Editing %				Base Editing %				Base Editing %			
	C9	C10	C18	C24	C9	C10	C18	C24	C9	C10	C18	C24
WT	0.04	0.05	0.46	3.59	0.05	0.05	4.63	1.37	4.87	4.65	0.08	0.04
E1370_Ala	0.04	0.05	0.56	4.16	0.04	0.04	2.97	0.63	0.52	0.42	0.04	0.02
E1370_Asn	0.07	0.05	0.96	5.61	0.06	0.04	3.86	1.24	1.73	1.6	0.03	0.03
E1370_Asp	0.04	0.04	0.29	2.73	0.03	0.04	1.83	0.36	2.1	1.99	0.05	0.02
E1370_Cys	0.04	0.04	0.47	3.62	0.04	0.03	2.79	0.71	3.36	2.79	0.06	0.04
E1370_Gln	0.05	0.04	0.59	3.72	0.04	0.04	2.44	0.51	3.71	3.25	0.05	0.04
E1370_Gly	0.07	0.05	1.13	5.88	0.07	0.05	4.63	1.27	4.01	3.68	0.08	0.03
E1370_His	0.08	0.06	1.47	8.27	0.06	0.05	4.06	1.5	5.12	4.56	0.06	0.03
E1370_Ile	0.04	0.04	0.13	1.51	0.04	0.05	0.81	0.16	0.86	0.46	0.03	0.03
E1370_Leu	0.04	0.03	0.25	2.27	0.03	0.04	1.59	0.27	2.69	2.09	0.05	0.03
E1370_Lys	0.04	0.04	0.37	3.51	0.05	0.04	2.78	0.63	2.34	1.51	0.05	0.02
E1370_Met	0.07	0.05	0.76	5.19	0.04	0.04	3.4	0.89	3.87	3.39	0.07	0.04
E1370_Phe	0.07	0.04	1.16	5.87	0.04	0.05	3.62	1.19	NA	NA	NA	NA
E1370_Pro	0.03	0.04	0.04	0.05	0.04	0.05	2.04	0.54	0.04	0.05	0.04	0.02
E1370_Ser	0.07	0.05	0.73	5.36	0.05	0.04	3.41	0.88	2.69	2.3	0.04	0.03
E1370_Thr	0.08	0.05	0.92	5.23	0.03	0.04	0.26	0.1	3.69	3.35	0.06	0.04
E1370_Trp	0.06	0.04	0.55	4.2	0.05	0.04	3.03	0.84	1.68	1.61	0.04	0.02
E1370_Tyr	0.08	0.05	0.94	4.76	0.03	0.04	3.65	0.96	0.88	0.77	0.04	0.02
E1370_Val	0.08	0.04	1.2	6.08	NA	NA	NA	NA	0.18	0.09	0.03	0.02

FIG. 6C

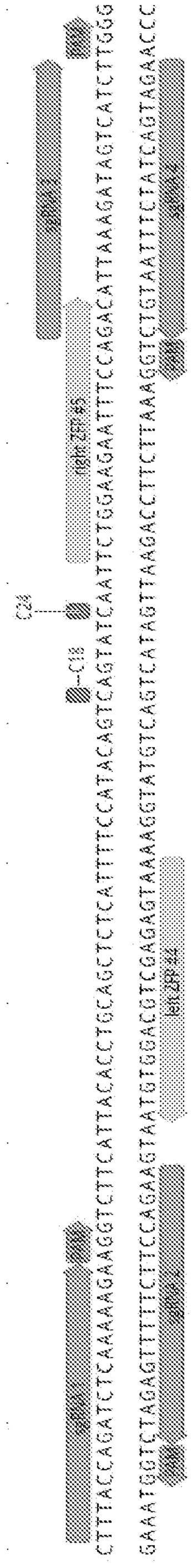

200
CAGTCAGTATCAATTCTGGAAGAATTTCCAGACAT
C24
C18
loxP SEQ #5
GTCAGTCATAGTTAAGACCTTCTTTAAAGGTCTGTA
CATTACACCTGCAGCTCTCATTTTCCATA
GTAATGTGGACGTCGAGAGTAAAAGGTAT
loxP SEQ #4
CTTTACCAGATCTCAAAAGAAGGTCTT
GAAATGGTCTAGAGTTTTTCTTCCAGAA
200

FIG. 7A

				Base Editing % (C>T)	
DddA	**Nickase**	**sgRNA**	**Nicked strand**	**C18**	**C24**
NA	NA	NA	NA	0.03	0.03
Left ZFP#4_G1333-N + Right ZFP#5_G1333-C (L26 linker)	NA	NA	NA	0.52	3.47
	SpCas_D10A	#1	Unedited	1.50	8.47
	SpCas_H840A	#1	Edited	1.07	6.23
	SpCas_D10A	#2	Edited	1.23	6.83
	SpCas_H840A	#2	Unedited	1.13	6.63
	SpCas_D10A	#3	Unedited	2.59	15.69
	SpCas_H840A	#3	Edited	1.05	6.11
	SpCas_D10A	#4	Edited	1.11	6.22
	SpCas_H840A	#4	Unedited	1.12	7.17

FIG. 7B

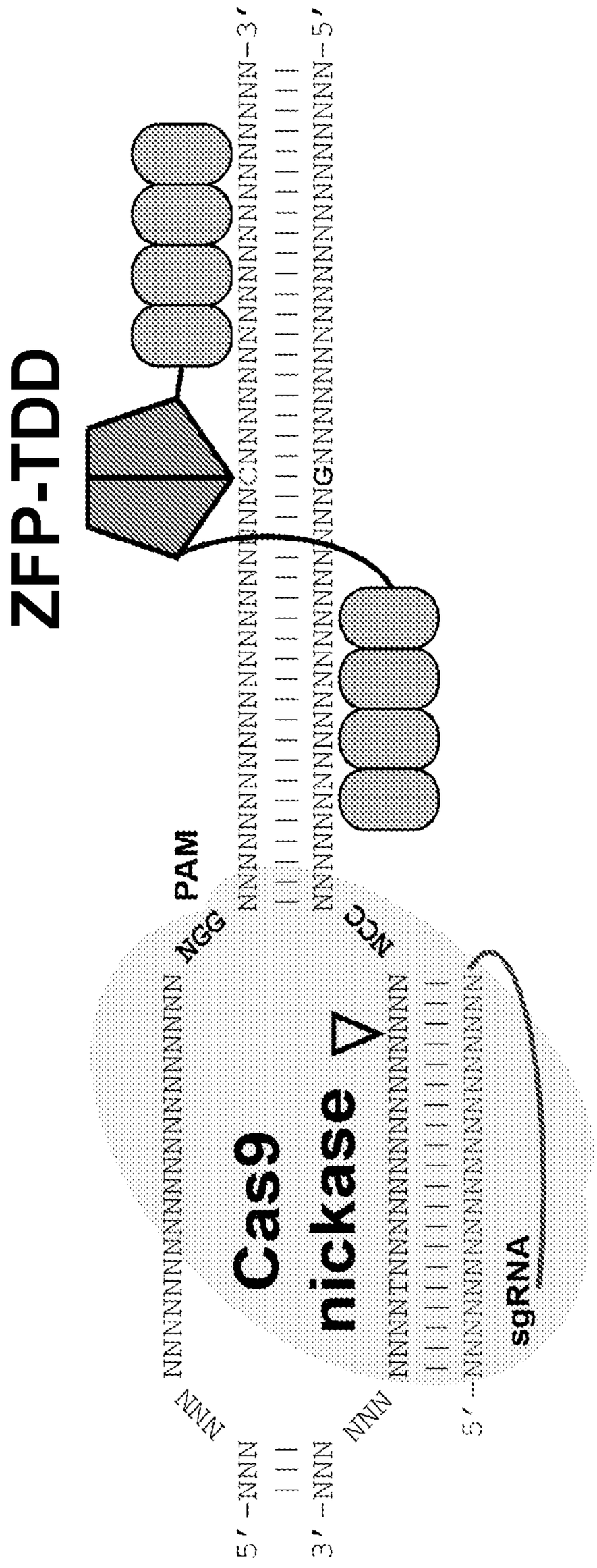

ZFP-TDD
Cas9
nickase
PAM
NGG
NCC
sgRNA
5'-NNNNNNNNNNNNNNNNNNNN-3'
3'-NNNNNNNNNNNNNNNNNNNN-5'

FIG. 8

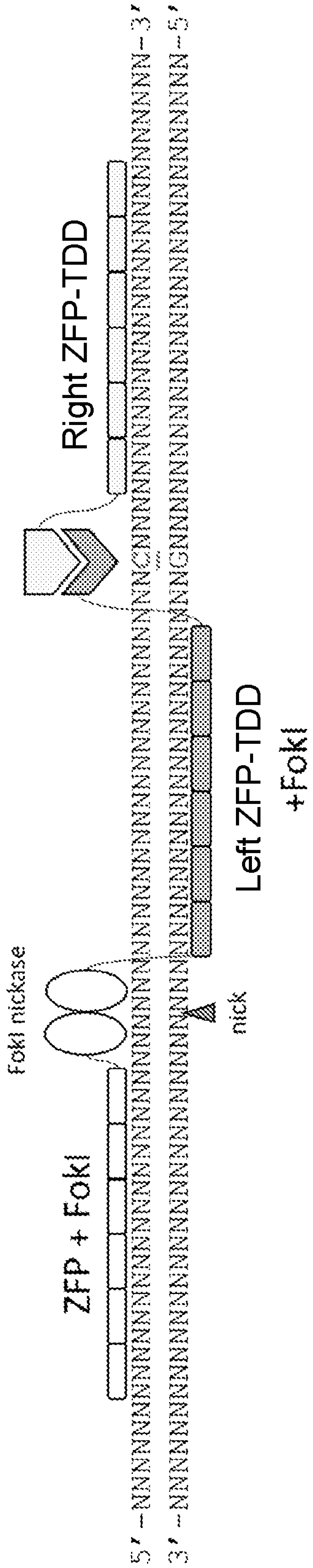

Right ZFP-TDD
Left ZFP-TDD
+FokI
ZFP + FokI
FokI nickase
nick
5'
3'
3'
5'

FIG. 9

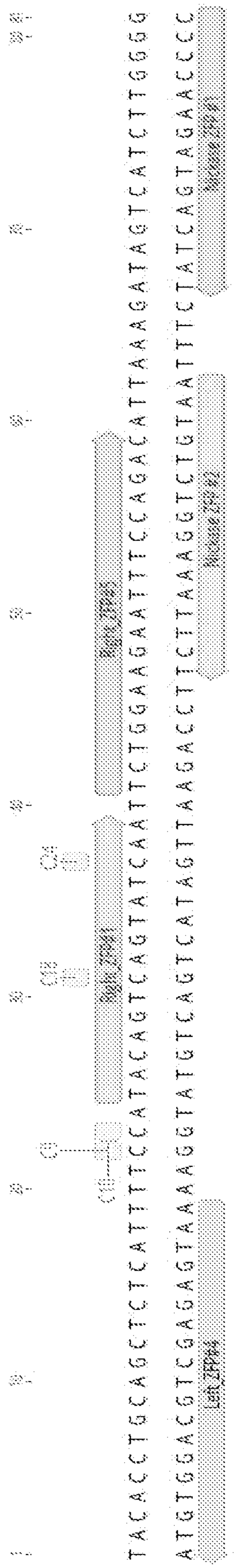


FIG. 10

			%Base Edits					
Left-DddA	**Right-DddA**	**Nick**	**C9**	**C10**	**C14**	**C18**	**C24**	**%Indels**
Left_ZFP#4-G1333-N	Right_ZFP#5-G1333-C	No nick	0.04	0.06	0.20	0.72	5.03	0.12
Left_ZFP#4-G1333-N	Right_ZFP#5-G1333-C	Cas9/sgRNA#3	0.09	0.04	0.75	2.62	18.13	0.92
Left_ZFP#4-G1333-N	FokI(ELD)-Right_ZFP#5-G1333-C	Nickase #1 (ZFP-FokI(KKR_F450N))	0.72	0.26	6.03	25.41	71.48	0.45
Left_ZFP#4-G1333-C	Right_ZFP#1-G1333-N	No nick	5.52	5.40	0.01	0.07	0.02	0.11
Left_ZFP#4-G1333-C	Right_ZFP#1-G1333-N	Cas9/sgRNA#3	40.85	39.51	0.04	0.25	0.10	9.57
Left_ZFP#4-G1333-C	FokI(ELD)-Right_ZFP#1-G1333-N	Nickase #2 (ZFP-FokI(KKR_F450N))	69.76	70.39	0.03	0.46	0.05	0.81

FIG. 11

				% Base Editing							
Deaminase	**Left construct**	**Right construct**	**Nickase**	**C9**	**C10**	**C14**	**C16**	**C18**	**C20**	**C24**	**Indel %**
DddA	Left_ZFP#4-G1333-N	FokI(ELD)-Right_ZFP#5-G1333-C		0.05	0.05	0.18	0.08	0.58	0.07	4.06	0.30
O1	Left_ZFP#4-G2278-N	FokI(ELD)-Right_ZFP#5-G2278-C	N/A	0.03	0.04	0.03	0.04	0.04	0.03	0.03	0.24
O1	Left_ZFP#4-G2278-N	FokI(ELD)-Right_ZFP#5-G2278-C	Nickase #1	0.03	0.04	0.03	0.04	0.04	0.03	0.04	0.25
O2	Left_ZFP#4-G1794-N	FokI(ELD)-Right_ZFP#5-G1794-C	N/A	0.03	0.04	0.05	0.04	0.04	0.03	0.05	0.23
O2	Left_ZFP#4-G1794-N	FokI(ELD)-Right_ZFP#5-G1794-C	Nickase #1	0.04	0.05	0.07	0.04	0.09	0.04	0.25	0.26
O3	Left_ZFP#4-G30-N	FokI(ELD)-Right_ZFP#5-G30-C	N/A	0.03	0.03	0.03	0.02	0.03	0.03	0.05	0.24
O3	Left_ZFP#4-G30-N	FokI(ELD)-Right_ZFP#5-G30-C	Nickase #1	0.02	0.03	0.03	0.04	0.03	0.04	0.04	0.26
O4	Left_ZFP#4-T161-N	FokI(ELD)-Right_ZFP#5-T161-C	N/A	0.03	0.04	0.04	0.03	0.03	0.03	0.03	0.25
O4	Left_ZFP#4-T161-N	FokI(ELD)-Right_ZFP#5-T161-C	Nickase #1	0.03	0.05	0.03	0.03	0.03	0.03	0.04	0.23
O5	Left_ZFP#4-G2299-N	FokI(ELD)-Right_ZFP#5-G2299-C	N/A	0.04	0.05	0.04	0.04	0.05	0.03	0.03	0.25
O5	Left_ZFP#4-G2299-N	FokI(ELD)-Right_ZFP#5-G2299-C	Nickase #1	0.04	0.05	0.03	0.03	0.06	0.03	0.06	0.25
O6	Left_ZFP#4-N2313-N	FokI(ELD)-Right_ZFP#5-N2313-C	N/A	0.11	0.06	1.19	0.28	0.39	0.87	0.90	0.26
O6	Left_ZFP#4-N2313-N	FokI(ELD)-Right_ZFP#5-N2313-C	Nickase #1	0.65	0.11	10.82	0.38	4.46	1.58	13.56	0.41
O1	Left_ZFP#4-G2278-C	FokI(ELD)-Right_ZFP#5-G2278-N	N/A	0.03	0.04	0.03	0.03	0.03	0.03	0.03	0.21
O1	Left_ZFP#4-G2278-C	FokI(ELD)-Right_ZFP#5-G2278-N	Nickase #1	0.03	0.03	0.03	0.03	0.03	0.03	0.04	0.25
O2	Left_ZFP#4-G1794-C	FokI(ELD)-Right_ZFP#5-G1794-N	N/A	0.04	0.05	0.03	0.05	0.07	0.03	0.04	0.24
O2	Left_ZFP#4-G1794-C	FokI(ELD)-Right_ZFP#5-G1794-N	Nickase #1	0.05	0.06	0.06	0.04	0.14	0.03	0.09	0.23
O3	Left_ZFP#4-G30-C	FokI(ELD)-Right_ZFP#5-G30-N	N/A	0.42	0.19	1.73	0.46	1.66	1.93	0.16	0.32
O3	Left_ZFP#4-G30-C	FokI(ELD)-Right_ZFP#5-G30-N	Nickase #1	1.16	0.34	6.19	1.25	4.87	2.90	0.74	0.36
O4	Left_ZFP#4-T161-C	FokI(ELD)-Right_ZFP#5-T161-N	N/A	0.03	0.04	0.07	0.04	0.07	0.06	0.03	0.24
O4	Left_ZFP#4-T161-C	FokI(ELD)-Right_ZFP#5-T161-N	Nickase #1	0.03	0.04	0.32	0.07	0.30	0.11	0.03	0.25
O5	Left_ZFP#4-G2299-C	FokI(ELD)-Right_ZFP#5-G2299-N	N/A	0.02	0.04	0.04	0.03	0.03	0.04	0.03	0.23
O5	Left_ZFP#4-G2299-C	FokI(ELD)-Right_ZFP#5-G2299-N	Nickase #1	0.04	0.05	0.04	0.03	0.05	0.03	0.04	0.25
O6	Left_ZFP#4-C2313-C	FokI(ELD)-Right_ZFP#5-N2313-N	N/A	0.05	0.04	0.36	0.05	0.26	0.35	0.04	0.24
O6	Left_ZFP#4-C2313-C	FokI(ELD)-Right_ZFP#5-N2313-N	Nickase #1	0.14	0.06	1.12	0.10	0.80	0.42	0.06	0.28
			EGFP control	0.04	0.04	0.04	0.03	0.04	0.04	0.04	0.27
			NTC control	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.12

FIG. 12

				% Base Editing							
Deaminase	Left construct	Right construct	Nickase	C9	C10	C14	C16	C18	C20	C24	Indel %
DddA	Left_ZFP#4-G1333-N	FokI(ELD)-Right_ZFP#1-G1333-C		3.32	3.18	0.04	0.06	0.06	0.08	0.04	0.32
O1	Left_ZFP#4-G2278-N	FokI(ELD)-Right_ZFP#1-G2278-C	N/A	0.03	0.05	0.03	0.03	0.04	0.03	0.03	0.23
O1	Left_ZFP#4-G2278-N	FokI(ELD)-Right_ZFP#1-G2278-C	Nickase #2	0.03	0.06	0.03	0.04	0.03	0.03	0.04	0.26
O2	Left_ZFP#4-G1794-N	FokI(ELD)-Right_ZFP#1-G1794-C	N/A	0.02	0.05	0.04	0.04	0.04	0.03	0.03	0.25
O2	Left_ZFP#4-G1794-N	FokI(ELD)-Right_ZFP#1-G1794-C	Nickase #2	0.03	0.04	0.03	0.03	0.05	0.03	0.03	0.25
O3	Left_ZFP#4-G30-N	FokI(ELD)-Right_ZFP#1-G30-C	N/A	0.94	0.15	0.68	0.15	0.28	0.58	0.14	0.27
O3	Left_ZFP#4-G30-N	FokI(ELD)-Right_ZFP#1-G30-C	Nickase #2	3.79	0.68	2.56	0.24	0.62	0.64	0.38	0.35
O4	Left_ZFP#4-T161-N	FokI(ELD)-Right_ZFP#1-T161-C	N/A	0.27	0.05	6.78	3.21	2.66	2.25	0.15	0.33
O4	Left_ZFP#4-T161-N	FokI(ELD)-Right_ZFP#1-T161-C	Nickase #2	0.57	0.08	15.77	3.89	5.67	2.53	0.45	0.40
O5	Left_ZFP#4-G2299-N	FokI(ELD)-Right_ZFP#1-G2299-C	N/A	0.03	0.05	0.03	0.04	0.02	0.03	0.04	0.26
O5	Left_ZFP#4-G2299-N	FokI(ELD)-Right_ZFP#1-G2299-C	Nickase #2	0.03	0.05	0.04	0.03	0.03	0.03	0.04	0.27
O6	Left_ZFP#4-N2313-N	FokI(ELD)-Right_ZFP#1-N2313-C	N/A	0.61	0.09	0.85	0.13	0.22	0.50	0.19	0.27
O6	Left_ZFP#4-N2313-N	FokI(ELD)-Right_ZFP#1-N2313-C	Nickase #2	2.42	0.22	5.38	0.13	0.68	0.46	0.82	0.33
O1	Left_ZFP#4-G2278-C	FokI(ELD)-Right_ZFP#1-G2278-N	N/A	0.03	0.04	0.03	0.04	0.03	0.04	0.03	0.25
O1	Left_ZFP#4-G2278-C	FokI(ELD)-Right_ZFP#1-G2278-N	Nickase #2	0.02	0.04	0.04	0.03	0.03	0.04	0.04	0.26
O2	Left_ZFP#4-G1794-C	FokI(ELD)-Right_ZFP#1-G1794-N	N/A	0.03	0.04	0.04	0.03	0.03	0.04	0.03	0.26
O2	Left_ZFP#4-G1794-C	FokI(ELD)-Right_ZFP#1-G1794-N	Nickase #2	0.03	0.05	0.04	0.03	0.05	0.04	0.04	0.27
O3	Left_ZFP#4-G30-C	FokI(ELD)-Right_ZFP#1-G30-N	N/A	0.17	0.06	0.05	0.05	0.04	0.19	0.02	0.26
O3	Left_ZFP#4-G30-C	FokI(ELD)-Right_ZFP#1-G30-N	Nickase #2	0.53	0.14	0.06	0.05	0.08	0.17	0.04	0.27
O4	Left_ZFP#4-T161-C	FokI(ELD)-Right_ZFP#1-T161-N	N/A	0.05	0.05	1.04	1.28	0.92	2.80	0.04	0.28
O4	Left_ZFP#4-T161-C	FokI(ELD)-Right_ZFP#1-T161-N	Nickase #2	0.09	0.06	1.72	1.06	1.26	2.53	0.06	0.30
O5	Left_ZFP#4-G2299-C	FokI(ELD)-Right_ZFP#1-G2299-N	N/A	0.03	0.04	0.03	0.03	0.03	0.03	0.03	0.23
O5	Left_ZFP#4-G2299-C	FokI(ELD)-Right_ZFP#1-G2299-N	Nickase #2	0.03	0.03	0.03	0.03	0.04	0.03	0.03	0.24
O6	Left_ZFP#4-C2313-C	FokI(ELD)-Right_ZFP#1-N2313-N	N/A	1.30	0.14	0.26	0.15	0.30	0.99	0.03	0.29
O6	Left_ZFP#4-C2313-C	FokI(ELD)-Right_ZFP#1-N2313-N	Nickase #2	7.12	0.77	0.80	0.17	0.58	1.09	0.14	0.42
			EGFP control	0.04	0.04	0.04	0.03	0.04	0.04	0.04	0.27
			NTC control	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.12

FIG. 12 (cont.)

Deaminase	Max %Base Editing
	No DNA nicking
DddA	4.11
O1	0.05
O2	0.06
O3	2.01
O4	7.29
O5	0.05
O6	1.38
EGFP	0.04

FIG. 13

Deaminase	#4-N/#5-C		#4-C/#5-N		#4-N/#1-C		#4-C/#1-N	
	Nickase		Nickase		Nickase		Nickase	
	No	Yes	No	Yes	No	Yes	No	Yes
DddA	4.11	ND	ND	ND	3.40	ND	ND	ND
O1	0.04	0.04	0.04	0.04	0.05	0.06	0.05	0.04
O2	0.05	0.25	0.06	0.14	0.05	0.05	0.05	0.05
O3	0.05	0.04	2.01	6.19	0.98	3.79	0.18	0.53
O4	0.05	0.05	0.08	0.32	7.29	15.77	2.78	2.53
O5	0.05	0.06	0.04	0.05	0.05	0.05	0.04	0.04
O6	1.12	13.56	0.34	1.12	0.84	5.38	1.38	7.12
EGFP	0.04	ND	ND	ND	0.04	ND	ND	ND

FIG. 14

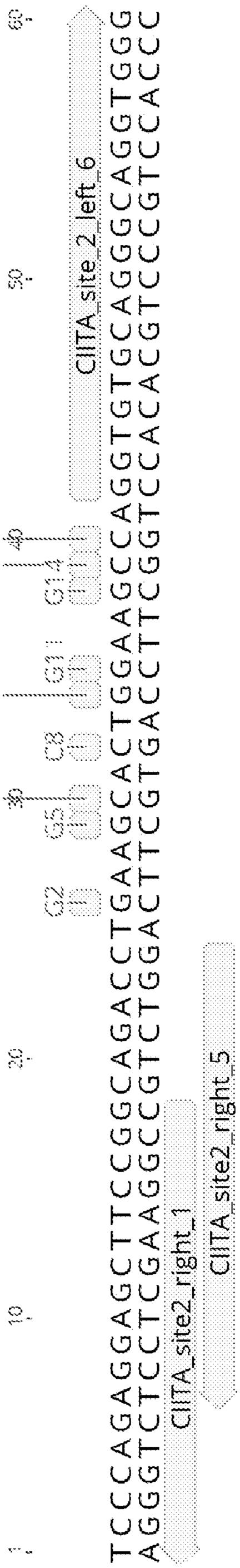

CTGGAAGCCAGGTGTGGGG
GACCTTCGGTCCACACCCC
CIITA_site2_left_6
G16
G15
40
G14
G10
G11
G8
CGGGCAGACCTGAAGCA
GCCCGTCTGGACTTCGT
CIITA_site2_right_5
G6
30
G5
G2
TCCCAGAGGAGCTTC
AGGGTCTCCTCGAAG
CIITA_site2_right_1
CIITA_site2_left_1

FIG. 15

	Highest %Base Editing CIITA site 2	
	Experiment #1	Experiment #2
EGFP	0.05	0.23
left_6_O14_G43N + right_5_O14_G43C	7.21	not tested
left_6_O14_A118N + right_5_O14_A118C	1.75	not tested
left_6_O1_S2346N + right_5_O1_S2346C	not tested	0.21
left_6_O1_G2278N + right_5_O1_G2278C	0.05	not tested
left_6_O8_T2175C + right_5_O8_T2175N	1.17	not tested
left_6_O8_G2108N + right_5_O8_G2108C	0.27	not tested
left_6_O5_S2367N + right_5_O5_S2367C	0.05	not tested
left_6_O5_G2299N + right_5_O5_G2299C	0.08	not tested
left_6_O9_P2179C + right_5_O9_P2179N	16.31	not tested
left_6_O9_G2112C + right_5_O9_G2112N	10.37	not tested
left_6_O17_G2156N + right_5_O17_G2156C	0.13	not tested
left_6_O17_G2087N + right_5_O17_G2087C	not tested	0.19
left_6_O4b_T161N + right_5_O4b_T161C	6.66	not tested
left_6_O4b_A229C + right_5_O4b_A229N	18.49	not tested
left_6_O4_T161C + right_5_O4_T161N	6.88	not tested
left_6_O4_A229N + right_5_O4_A229C	20.65	not tested
left_6_O6b_R2385N + right_5_O6b_R2385C	31.79	not tested
left_6_O6b_N2313N + right_5_O6b_N2313C	18.16	not tested
left_6_O6_R2385C + right_5_O6_R2385N	45.50	not tested
left_6_O6_N2313C + right_5_O6_N2313N	10.86	not tested

FIG. 16

	Highest %Base Editing CIITA site 2	
	Experiment #1	Experiment #2
left_6_O2b_P1861C + right_5_O2b_P1861N	0.07	not tested
left_6_O2b_G1794N + right_5_O2b_G1794C	0.08	not tested
left_6_O2_P1861N + right_5_O2_P1861C	0.09	not tested
left_6_O2_G1794C + right_5_O2_G1794N	0.07	not tested
left_6_O7_G33N + right_5_O7_G33C	0.06	not tested
left_6_O7_G102N + right_5_O7_G102C	0.07	not tested
left_6_O13_G1411N + right_5_O13_G1411C	0.07	not tested
left_6_O13_A1477C + right_5_O13_A1477N	0.06	not tested
left_6_O3_N94C + right_5_O3_N94N	not tested	6.10
left_6_O3_G30C + right_5_O3_G30N	not tested	9.48
left_6_O10_G1746C + right_5_O10_G1746N	2.35	not tested
left_6_O10_G1667C + right_5_O10_G1667N	0.06	not tested
left_6_O15_G315N + right_5_O15_G315C	4.50	not tested
left_6_O15_A390N + right_5_O15_A390C	5.10	not tested
left_6_O16_G1342N + right_5_O16_G1342C	not tested	0.20
left_6_O16_G1264N + right_5_O16_G1264C	0.08	not tested
left_6_O19_G321N + right_1_O19_G321C	not tested	9.15
left_6_O19_G234N + right_5_O19_G234C	0.05	not tested
left_6_O18_G181N + right_5_O18_G181C	2.21	not tested
left_6_O18_A250N + right_5_O18_A250C	6.16	not tested
left_6_O11_G1502N + right_5_O11_G1502C	0.08	not tested
left_6_O11_G1430C + right_5_O11_G1430N	0.12	not tested
left_6_O11_A1498C + right_5_O11_A1498N	0.07	not tested
left_6_O12_G1421N + right_5_O12_G1421C	0.06	not tested
left_6_O12_A1488N + right_5_O12_A1488C	0.57	not tested

FIG. 16 (cont.)

		%Base Editing							
Amplicon Position	Motif	DddA	O18	O15	O10	O4	O6	O9	O14
52	GC**C**	NA	0.05	0.04	0.04	0.11	0.07	0.06	0.05
53	TG**C**	NA	0.09	0.11	0.12	0.12	0.17	0.13	0.08
54	GG**C**	NA	0.06	0.14	0.07	0.11	0.11	0.24	0.10
56	GT**C**	NA	0.37	0.51	0.92	19.39	24.53	17.63	0.20
58	GA**C**	1.12	2.25	0.94	0.20	7.22	1.94	3.14	0.62
59	AC**C**	0.70	0.38	0.24	0.47	1.10	1.33	0.92	0.19
61	TT**C**	4.74	0.89	1.63	0.10	3.24	16.88	7.85	1.57
64	TG**C**	0.02	0.06	0.11	0.04	4.34	20.27	2.52	0.08
65	AG**C**	0.34	0.17	0.38	0.12	6.45	24.42	5.72	0.52
67	CA**C**	0.51	0.92	1.23	1.94	20.37	46.21	18.20	0.69
69	TC**C**	17.89	4.59	1.66	1.24	21.31	29.95	2.81	2.73
70	TT**C**	29.93	6.15	6.09	2.30	21.09	36.57	13.70	8.39
73	GG**C**	0.02	0.03	0.05	0.03	0.74	0.31	0.26	0.05
74	AG**C**	0.10	0.08	0.07	0.08	0.19	4.17	0.31	0.06
75	GC**C**	0.22	0.04	0.05	0.05	0.09	0.67	0.05	0.03
77	AC**C**	0.04	0.12	0.03	0.05	0.17	0.24	0.10	0.03
78	CA**C**	0.04	0.26	0.03	0.04	1.41	0.31	0.20	0.04

Amplicon:

CTGGGGCAGCTGATCACATGTTTTCTCTGCAGCCTTCCCAGAGGAGCTTCCGGCAGACCTGAAGCACTGG
AAGCCAGGTGTGCAGGGCAGGTGGGCTGGGGTTGGGAAGGGTGGATGCCTTGGGGAGGGGATGGAAG

FIG. 17

	%Base Editing								
	G2	G5	C6	C8	**G10**	**G11**	G14	C15	C16
left_6_O6_R2385C_L26 + right_5_O6_R2385N_L26	14.37	18.66	25.92	47.30	37.58	44.94	0.11	0.29	0.15
left_6_O6_R2385C_L21 + right_5_O6_R2385N_L26	10.74	9.55	18.53	46.88	42.54	48.62	0.10	0.31	0.22
left_6_O6_R2385C_L18 + right_5_O6_R2385N_L26	4.90	2.01	7.18	46.23	45.00	50.44	0.06	0.32	0.14
left_6_O6_R2385C_L13 + right_5_O6_R2385N_L26	1.02	1.03	3.37	35.91	50.22	51.56	0.07	0.18	0.13
left_6_O6_R2385C_L11 + right_5_O6_R2385N_L26	0.23	0.16	1.78	19.10	35.28	36.01	0.03	0.10	0.12
left_6_O6_R2385C_L9 + right_5_O6_R2385N_L26	0.11	0.09	2.55	6.75	28.61	30.30	0.03	0.07	0.05
NA + right_5_O6_R2385N_L26	NA	NA	NA	NA	NA	NA	NA	NA	NA
left_6_O6_R2385C_L4 + right_5_O6_R2385N_L26	0.06	0.07	1.95	1.75	10.63	12.20	0.02	0.06	0.03
left_6_O6_R2385C_L26 + right_5_O6_R2385N_L21	14.19	15.81	21.90	44.54	32.49	39.82	0.08	0.28	0.09
left_6_O6_R2385C_L21 + right_5_O6_R2385N_L21	11.09	8.76	17.46	47.21	40.18	46.53	0.06	0.27	0.12
left_6_O6_R2385C_L18 + right_5_O6_R2385N_L21	4.54	1.66	6.15	41.46	38.83	44.03	0.03	0.24	0.09
left_6_O6_R2385C_L13 + right_5_O6_R2385N_L21	1.03	1.07	3.48	36.49	48.03	49.19	0.04	0.17	0.17
left_6_O6_R2385C_L11 + right_5_O6_R2385N_L21	0.23	0.25	1.79	18.04	33.05	33.71	0.05	0.10	0.11
left_6_O6_R2385C_L9 + right_5_O6_R2385N_L21	0.13	0.05	2.10	6.47	26.39	27.98	0.03	0.06	0.05
NA + right_5_O6_R2385N_L21	NA	NA	NA	NA	NA	NA	NA	NA	NA
left_6_O6_R2385C_L4 + right_5_O6_R2385N_L21	0.08	0.05	1.60	1.35	8.61	9.98	0.03	0.07	0.05
left_6_O6_R2385C_L26 + right_5_O6_R2385N_L18	14.85	16.67	23.20	45.87	33.67	40.58	0.04	0.23	0.11
left_6_O6_R2385C_L21 + right_5_O6_R2385N_L18	10.48	7.94	16.97	45.14	38.82	44.76	0.07	0.25	0.07
left_6_O6_R2385C_L18 + right_5_O6_R2385N_L18	3.99	1.60	6.42	39.94	38.15	42.69	0.06	0.19	0.11
left_6_O6_R2385C_L13 + right_5_O6_R2385N_L18	1.12	1.03	3.41	35.01	45.51	46.95	0.05	0.17	0.07
left_6_O6_R2385C_L11 + right_5_O6_R2385N_L18	0.25	0.26	2.21	20.19	34.47	35.10	0.02	0.13	0.1
left_6_O6_R2385C_L9 + right_5_O6_R2385N_L18	0.13	0.09	1.83	7.86	26.63	28.13	0.02	0.03	0.04
NA + right_5_O6_R2385N_L18	NA	NA	NA	NA	NA	NA	NA	NA	NA
left_6_O6_R2385C_L4 + right_5_O6_R2385N_L18	0.07	0.08	1.48	2.26	9.16	10.41	0.03	0.06	0.05

FIG. 18

	%Base Editing								
	G2	G5	C6	C8	**G10**	**G11**	G14	C15	C16
left_6_O6_R2385C_L26 + right_5_O6_R2385N_L13	14.04	15.29	23.43	44.66	33.03	39.52	0.04	0.20	0.04
left_6_O6_R2385C_L21 + right_5_O6_R2385N_L13	10.17	7.58	17.97	44.93	38.35	43.91	0.07	0.19	0.05
left_6_O6_R2385C_L18 + right_5_O6_R2385N_L13	4.03	1.67	7.72	42.03	39.26	43.78	0.03	0.19	0.04
left_6_O6_R2385C_L13 + right_5_O6_R2385N_L13	1.05	1.04	4.22	36.59	47.40	48.60	0.04	0.11	0.04
left_6_O6_R2385C_L11 + right_5_O6_R2385N_L13	0.28	0.23	1.81	20.17	34.36	35.04	0.03	0.08	0.08
left_6_O6_R2385C_L9 + right_5_O6_R2385N_L13	0.09	0.07	2.04	6.85	27.25	28.70	0.03	0.07	0.03
NA + right_5_O6_R2385N_L13	NA	NA	NA	NA	NA	NA	NA	NA	NA
left_6_O6_R2385C_L4 + right_5_O6_R2385N_L13	0.06	0.05	1.36	1.40	7.30	8.54	0.04	0.04	0.04
left_6_O6_R2385C_L26 + right_5_O6_R2385N_L11	12.88	14.63	24.07	43.60	31.16	38.02	0.04	0.20	0.04
left_6_O6_R2385C_L21 + right_5_O6_R2385N_L11	9.51	6.81	18.13	43.42	35.85	41.59	0.04	0.17	0.04
left_6_O6_R2385C_L18 + right_5_O6_R2385N_L11	3.92	1.87	9.04	41.95	39.11	44.16	0.06	0.18	0.03
left_6_O6_R2385C_L13 + right_5_O6_R2385N_L11	0.98	0.95	4.24	36.59	46.16	47.62	0.02	0.12	0.05
left_6_O6_R2385C_L11 + right_5_O6_R2385N_L11	0.18	0.24	1.70	18.94	31.90	32.39	0.03	0.11	0.06
left_6_O6_R2385C_L9 + right_5_O6_R2385N_L11	0.10	0.06	1.25	5.46	23.11	24.45	0.01	0.08	0.03
NA + right_5_O6_R2385N_L11	NA	NA	NA	NA	NA	NA	NA	NA	NA
left_6_O6_R2385C_L4 + right_5_O6_R2385N_L11	0.07	0.05	0.93	1.15	7.13	8.39	0.03	0.05	0.04
left_6_O6_R2385C_L26 + right_5_O6_R2385N_L9	11.42	15.44	27.83	44.78	33.19	40.13	0.03	0.17	0.03
left_6_O6_R2385C_L21 + right_5_O6_R2385N_L9	7.60	6.78	20.26	43.18	36.70	42.30	0.02	0.20	0.04
left_6_O6_R2385C_L18 + right_5_O6_R2385N_L9	2.23	1.36	7.28	29.83	29.85	33.21	0.03	0.11	0.04
left_6_O6_R2385C_L13 + right_5_O6_R2385N_L9	0.85	1.02	4.69	35.32	46.86	47.98	0.04	0.13	0.07
left_6_O6_R2385C_L11 + right_5_O6_R2385N_L9	0.16	0.26	2.08	17.22	32.31	32.95	0.04	0.07	0.05
left_6_O6_R2385C_L9 + right_5_O6_R2385N_L9	0.09	0.08	1.56	5.14	24.14	25.41	0.03	0.04	0.03
NA + right_5_O6_R2385N_L9	NA	NA	NA	NA	NA	NA	NA	NA	NA
left_6_O6_R2385C_L4 + right_5_O6_R2385N_L9	0.06	0.06	1.27	1.61	7.77	8.73	0.02	0.07	0.04

FIG. 18 (cont.)

	%Base Editing								
	G2	G5	C6	C8	**G10**	**G11**	G14	C15	C16
left_6_O6_R2385C_L26 + right_5_O6_R2385N_L6	9.42	12.64	21.16	39.22	26.83	33.41	0.06	0.11	0.05
left_6_O6_R2385C_L21 + right_5_O6_R2385N_L6	5.71	5.44	15.63	39.10	31.47	37.16	0.03	0.16	0.08
left_6_O6_R2385C_L18 + right_5_O6_R2385N_L6	2.41	1.14	8.30	38.48	34.77	39.95	0.06	0.13	0.07
left_6_O6_R2385C_L13 + right_5_O6_R2385N_L6	0.66	0.74	4.56	33.09	41.88	43.10	0.04	0.10	0.06
left_6_O6_R2385C_L11 + right_5_O6_R2385N_L6	0.10	0.09	1.53	12.72	25.51	26.13	0.02	0.08	0.04
left_6_O6_R2385C_L9 + right_5_O6_R2385N_L6	0.05	0.05	1.29	4.18	20.43	21.81	0.04	0.05	0.03
NA + right_5_O6_R2385N_L6	NA	NA	NA	NA	NA	NA	NA	NA	NA
left_6_O6_R2385C_L4 + right_5_O6_R2385N_L6	0.05	0.05	1.03	1.24	5.81	6.97	0.03	0.03	0.03
left_6_O6_R2385C_L26 + right_5_O6_R2385N_L4	6.26	8.74	15.02	34.33	22.63	28.85	0.04	0.09	0.04
left_6_O6_R2385C_L21 + right_5_O6_R2385N_L4	4.83	4.33	11.99	36.34	29.10	34.84	0.06	0.15	0.04
left_6_O6_R2385C_L18 + right_5_O6_R2385N_L4	1.88	1.01	5.71	36.03	31.98	37.04	0.05	0.12	0.03
left_6_O6_R2385C_L13 + right_5_O6_R2385N_L4	0.50	0.48	3.37	27.94	36.61	37.80	0.05	0.11	0.05
left_6_O6_R2385C_L11 + right_5_O6_R2385N_L4	0.13	0.13	1.66	10.21	21.04	21.81	0.03	0.10	0.05
left_6_O6_R2385C_L9 + right_5_O6_R2385N_L4	0.08	0.04	1.41	3.76	18.81	20.14	0.03	0.05	0.03
NA + right_5_O6_R2385N_L4	NA	NA	NA	NA	NA	NA	NA	NA	NA
left_6_O6_R2385C_L4 + right_5_O6_R2385N_L4	0.03	0.04	0.95	0.98	4.97	5.99	0.03	0.07	0.03

FIG. 18 (cont.)

	%Base Editing								
	G2	G5	C6	C8	**G10**	**G11**	G14	C15	C16
left_6_O6_R2385N_L26 + right_5_O6_R2385C_L26	5.90	0.36	3.70	25.98	39.04	41.70	0.16	3.08	0.04
left_6_O6_R2385N_L21 + right_5_O6_R2385C_L26	4.15	0.10	3.09	20.93	35.53	37.83	0.15	3.65	0.47
left_6_O6_R2385N_L18 + right_5_O6_R2385C_L26	3.28	0.08	2.68	17.20	32.23	34.41	0.09	4.04	0.67
left_6_O6_R2385N_L13 + right_5_O6_R2385C_L26	1.04	0.06	1.77	11.44	26.48	28.59	0.09	4.80	0.67
left_6_O6_R2385N_L11 + right_5_O6_R2385C_L26	0.60	0.06	1.69	9.44	25.21	27.36	0.08	6.75	0.92
left_6_O6_R2385N_L9 + right_5_O6_R2385C_L26	0.25	0.06	1.76	7.00	24.51	26.72	0.05	8.16	0.84
left_6_O6_R2385N_L6 + right_5_O6_R2385C_L26	0.14	0.06	2.03	6.41	25.02	28.26	0.06	5.91	0.73
left_6_O6_R2385N_L4 + right_5_O6_R2385C_L26	0.07	0.05	1.57	4.91	21.32	25.08	0.05	5.29	0.49
left_6_O6_R2385N_L26 + right_5_O6_R2385C_L21	7.44	0.49	4.82	29.61	42.21	44.89	0.27	3.16	0.31
left_6_O6_R2385N_L21 + right_5_O6_R2385C_L21	5.34	0.16	4.01	24.93	39.82	42.32	0.13	3.80	0.44
left_6_O6_R2385N_L18 + right_5_O6_R2385C_L21	3.71	0.11	3.29	18.90	31.64	33.73	0.10	3.82	0.69
left_6_O6_R2385N_L13 + right_5_O6_R2385C_L21	1.36	0.11	2.02	14.69	27.23	29.55	0.09	4.64	0.69
left_6_O6_R2385N_L11 + right_5_O6_R2385C_L21	0.62	0.06	1.80	11.43	24.59	26.64	0.05	6.36	0.72
left_6_O6_R2385N_L9 + right_5_O6_R2385C_L21	0.26	0.04	1.91	8.99	24.48	26.54	0.05	8.10	0.83
left_6_O6_R2385N_L6 + right_5_O6_R2385C_L21	0.17	0.04	2.18	7.65	24.92	27.82	0.05	5.31	0.68
left_6_O6_R2385N_L4 + right_5_O6_R2385C_L21	0.12	0.04	1.74	6.14	23.13	26.60	0.04	5.16	0.38
left_6_O6_R2385N_L26 + right_5_O6_R2385C_L18	9.66	0.64	4.63	32.55	39.82	43.94	0.18	3.21	0.34
left_6_O6_R2385N_L21 + right_5_O6_R2385C_L18	7.02	0.27	3.87	27.81	39.20	42.93	0.15	4.14	0.54
left_6_O6_R2385N_L18 + right_5_O6_R2385C_L18	5.18	0.13	2.98	22.88	33.25	36.49	0.11	4.23	0.64
left_6_O6_R2385N_L13 + right_5_O6_R2385C_L18	1.89	0.11	2.05	16.13	26.51	29.52	0.04	5.53	0.67
left_6_O6_R2385N_L11 + right_5_O6_R2385C_L18	0.86	0.11	1.93	13.94	24.98	27.55	0.09	6.92	0.63
left_6_O6_R2385N_L9 + right_5_O6_R2385C_L18	0.37	0.09	1.46	10.44	24.16	27.25	0.07	8.28	0.66
left_6_O6_R2385N_L6 + right_5_O6_R2385C_L18	0.16	0.08	1.59	8.25	23.14	26.37	0.05	5.43	0.58
left_6_O6_R2385N_L4 + right_5_O6_R2385C_L18	0.10	0.04	1.52	7.25	22.19	26.04	0.03	5.22	0.29

FIG. 18 (cont.)

	%Base Editing								
	G2	G5	C6	C8	**G10**	**G11**	G14	C15	C16
left_6_O6_R2385N_L26 + right_5_O6_R2385C_L13	22.58	1.31	7.47	39.96	40.63	45.13	0.42	5.52	0.21
left_6_O6_R2385N_L21 + right_5_O6_R2385C_L13	17.55	0.53	5.25	35.67	40.57	44.51	0.28	7.17	1.99
left_6_O6_R2385N_L18 + right_5_O6_R2385C_L13	12.31	0.25	3.85	28.03	33.91	37.26	0.16	6.75	2.59
left_6_O6_R2385N_L13 + right_5_O6_R2385C_L13	3.79	0.15	2.59	19.23	24.16	27.09	0.10	8.89	2.64
left_6_O6_R2385N_L11 + right_5_O6_R2385C_L13	2.00	0.19	2.34	17.80	25.52	28.14	0.10	13.32	3.31
left_6_O6_R2385N_L9 + right_5_O6_R2385C_L13	0.82	0.19	1.83	13.60	24.37	27.02	0.07	14.52	4.19
left_6_O6_R2385N_L6 + right_5_O6_R2385C_L13	0.35	0.12	1.73	12.32	25.24	28.29	0.08	9.80	3.5
left_6_O6_R2385N_L4 + right_5_O6_R2385C_L13	0.26	0.07	1.36	10.35	23.33	26.77	0.08	9.41	1.57
left_6_O6_R2385N_L26 + right_5_O6_R2385C_L11	28.12	1.17	5.69	41.75	33.31	38.97	0.26	5.50	1.19
left_6_O6_R2385N_L21 + right_5_O6_R2385C_L11	22.74	0.48	4.53	38.71	33.07	38.08	0.30	7.06	1.8
left_6_O6_R2385N_L18 + right_5_O6_R2385C_L11	15.85	0.21	3.27	32.04	28.33	32.54	0.15	6.90	2.05
left_6_O6_R2385N_L13 + right_5_O6_R2385C_L11	5.63	0.16	2.79	24.19	25.26	29.39	0.10	12.11	2.01
left_6_O6_R2385N_L11 + right_5_O6_R2385C_L11	2.69	0.23	2.28	20.70	24.11	27.77	0.07	17.54	3.6
left_6_O6_R2385N_L9 + right_5_O6_R2385C_L11	1.04	0.22	1.91	15.39	22.56	26.07	0.07	17.98	4.91
left_6_O6_R2385N_L6 + right_5_O6_R2385C_L11	0.48	0.17	1.55	13.73	23.34	26.47	0.08	11.81	3.67
left_6_O6_R2385N_L4 + right_5_O6_R2385C_L11	0.30	0.11	1.13	11.69	19.94	24.09	0.04	10.40	1.58
left_6_O6_R2385N_L26 + right_5_O6_R2385C_L9	19.81	0.60	2.65	34.29	19.59	25.98	0.08	2.30	0.92
left_6_O6_R2385N_L21 + right_5_O6_R2385C_L9	15.49	0.26	2.24	31.89	20.21	25.91	0.10	3.21	0.24
left_6_O6_R2385N_L18 + right_5_O6_R2385C_L9	10.53	0.16	1.84	24.95	17.51	22.14	0.09	3.67	0.32
left_6_O6_R2385N_L13 + right_5_O6_R2385C_L9	3.17	0.09	1.41	16.78	15.19	19.40	0.05	6.31	0.41
left_6_O6_R2385N_L11 + right_5_O6_R2385C_L9	1.53	0.09	1.36	14.34	16.27	21.16	0.05	12.26	0.87
left_6_O6_R2385N_L9 + right_5_O6_R2385C_L9	0.65	0.10	1.18	12.41	17.57	22.69	0.04	15.80	1.59
left_6_O6_R2385N_L6 + right_5_O6_R2385C_L9	0.18	0.10	0.95	9.41	15.53	21.13	0.03	9.87	1.35
left_6_O6_R2385N_L4 + right_5_O6_R2385C_L9	0.14	0.06	0.59	7.84	14.72	21.10	0.05	9.95	0.64

FIG. 18 (cont.)

	%Base Editing								
	G2	G5	C6	C8	**G10**	**G11**	G14	C15	C16
left_6_O6_R2385N_L26 + right_5_O6_R2385C_L6	20.77	1.29	5.30	29.89	18.80	25.61	0.26	3.55	0.35
left_6_O6_R2385N_L21 + right_5_O6_R2385C_L6	13.52	0.40	4.07	25.98	17.10	22.71	0.12	4.21	0.44
left_6_O6_R2385N_L18 + right_5_O6_R2385C_L6	9.88	0.23	3.42	19.74	15.16	19.31	0.11	4.47	0.7
left_6_O6_R2385N_L13 + right_5_O6_R2385C_L6	3.66	0.16	2.72	14.08	13.88	18.24	0.03	7.93	0.53
left_6_O6_R2385N_L11 + right_5_O6_R2385C_L6	1.94	0.15	2.13	11.76	13.31	17.38	0.04	11.43	0.99
left_6_O6_R2385N_L9 + right_5_O6_R2385C_L6	0.86	0.16	1.72	9.42	14.93	19.54	0.05	13.98	1.22
left_6_O6_R2385N_L6 + right_5_O6_R2385C_L6	0.23	0.09	1.40	7.82	14.13	19.75	0.04	9.73	0.98
left_6_O6_R2385N_L4 + right_5_O6_R2385C_L6	0.08	0.16	1.07	6.54	13.71	19.80	0.03	9.13	0.46
left_6_O6_R2385N_L26 + NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
left_6_O6_R2385N_L21 + NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
left_6_O6_R2385N_L18 + NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
left_6_O6_R2385N_L13 + NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
left_6_O6_R2385N_L11 + NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
left_6_O6_R2385N_L9 + NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
left_6_O6_R2385N_L6 + NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
left_6_O6_R2385N_L4 + NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

FIG. 18 (cont.)

	%Base Editing								
Reagent	G2	G5	C6	C8	**G10**	**G11**	G14	C15	C16
s2_left_6_O14_G43N_L26+s2_right_5_O14_G43C_L26	0.74	0.05	0.16	0.27	1.02	9.32	0.10	0.06	0.03
s2_left_6_O14_G43N_L21+s2_right_5_O14_G43C_L26	0.50	0.04	0.13	0.28	0.70	9.34	0.08	0.04	0.04
s2_left_6_O14_G43N_L18+s2_right_5_O14_G43C_L26	0.38	0.05	0.13	0.19	0.83	9.56	0.10	0.03	0.02
s2_left_6_O14_G43N_L13+s2_right_5_O14_G43C_L26	0.15	0.03	0.11	0.21	0.48	8.12	0.08	0.05	0.01
s2_left_6_O14_G43N_L11+s2_right_5_O14_G43C_L26	0.11	0.05	0.15	0.27	0.71	11.26	0.06	0.03	0.03
s2_left_6_O14_G43N_L9+s2_right_5_O14_G43C_L26	0.08	0.02	0.11	0.17	0.55	11.20	0.08	0.05	0.02
s2_left_6_O14_G43N_L6+s2_right_5_O14_G43C_L26	0.05	0.04	0.12	0.18	0.43	9.37	0.09	0.06	0.03
s2_left_6_O14_G43N_L4+s2_right_5_O14_G43C_L26	0.04	0.04	0.09	0.18	0.32	8.31	0.08	0.05	0.03
s2_left_6_O14_G43N_L26+s2_right_5_O14_G43C_L21	0.66	0.07	0.12	0.31	0.87	9.91	0.10	0.06	0.02
s2_left_6_O14_G43N_L21+s2_right_5_O14_G43C_L21	0.49	0.07	0.15	0.25	0.83	10.58	0.10	0.05	0.02
s2_left_6_O14_G43N_L18+s2_right_5_O14_G43C_L21	0.40	0.08	0.14	0.35	0.82	9.89	0.07	0.03	0.02
s2_left_6_O14_G43N_L13+s2_right_5_O14_G43C_L21	0.22	0.05	0.17	0.23	0.49	8.79	0.09	0.08	0.03
s2_left_6_O14_G43N_L11+s2_right_5_O14_G43C_L21	0.14	0.04	0.13	0.24	0.58	10.81	0.05	0.06	0.01
s2_left_6_O14_G43N_L9+s2_right_5_O14_G43C_L21	0.08	0.05	0.13	0.22	0.68	12.06	0.07	0.05	0.03
s2_left_6_O14_G43N_L6+s2_right_5_O14_G43C_L21	0.07	0.05	0.12	0.18	0.50	10.70	0.09	0.06	0.03
s2_left_6_O14_G43N_L4+s2_right_5_O14_G43C_L21	0.06	0.05	0.08	0.21	0.32	8.56	0.05	0.07	0.03
s2_left_6_O14_G43N_L26+s2_right_5_O14_G43C_L18	0.75	0.09	0.20	0.37	0.81	9.64	0.08	0.05	0.03
s2_left_6_O14_G43N_L21+s2_right_5_O14_G43C_L18	0.64	0.08	0.19	0.34	0.93	11.15	0.14	0.04	0.03
s2_left_6_O14_G43N_L18+s2_right_5_O14_G43C_L18	0.55	0.08	0.23	0.37	0.92	11.09	0.11	0.08	0.03
s2_left_6_O14_G43N_L13+s2_right_5_O14_G43C_L18	0.18	0.05	0.13	0.35	0.49	8.85	0.10	0.05	0.02
s2_left_6_O14_G43N_L11+s2_right_5_O14_G43C_L18	0.14	0.09	0.16	0.28	0.64	10.90	0.09	0.06	0.04
s2_left_6_O14_G43N_L9+s2_right_5_O14_G43C_L18	0.07	0.03	0.18	0.19	0.65	11.43	0.08	0.05	0.03
s2_left_6_O14_G43N_L6+s2_right_5_O14_G43C_L18	0.05	0.05	0.11	0.18	0.38	8.95	0.07	0.07	0.03
s2_left_6_O14_G43N_L4+s2_right_5_O14_G43C_L18	0.03	0.05	0.11	0.16	0.33	8.92	0.05	0.05	0.03

FIG. 18 (cont.)

	%Base Editing								
	G2	G5	C6	C8	**G10**	**G11**	G14	C15	C16
s2_left_6_O14_G43N_L26+s2_right_5_O14_G43C_L13	0.79	0.15	0.14	0.34	1.21	9.97	0.07	0.07	0.02
s2_left_6_O14_G43N_L21+s2_right_5_O14_G43C_L13	0.50	0.11	0.13	0.34	0.86	9.53	0.09	0.09	0.03
s2_left_6_O14_G43N_L18+s2_right_5_O14_G43C_L13	0.54	0.12	0.20	0.30	0.95	10.24	0.07	0.03	0.03
s2_left_6_O14_G43N_L13+s2_right_5_O14_G43C_L13	0.18	0.05	0.14	0.32	0.58	9.03	0.06	0.07	0.04
s2_left_6_O14_G43N_L11+s2_right_5_O14_G43C_L13	0.15	0.09	0.14	0.18	0.73	10.57	0.07	0.06	0.03
s2_left_6_O14_G43N_L9+s2_right_5_O14_G43C_L13	0.10	0.04	0.13	0.17	0.72	11.73	0.06	0.05	0.03
s2_left_6_O14_G43N_L6+s2_right_5_O14_G43C_L13	0.06	0.07	0.15	0.16	0.55	10.05	0.09	0.05	0.04
s2_left_6_O14_G43N_L4+s2_right_5_O14_G43C_L13	0.06	0.07	0.12	0.12	0.35	8.52	0.07	0.05	0.02
s2_left_6_O14_G43N_L26+s2_right_5_O14_G43C_L11	0.85	0.26	0.14	0.35	1.38	10.46	0.07	0.07	0.03
s2_left_6_O14_G43N_L21+s2_right_5_O14_G43C_L11	0.55	0.12	0.11	0.34	0.91	9.22	0.03	0.10	0.01
s2_left_6_O14_G43N_L18+s2_right_5_O14_G43C_L11	0.49	0.14	0.14	0.40	1.22	11.37	0.05	0.06	0.02
s2_left_6_O14_G43N_L13+s2_right_5_O14_G43C_L11	0.19	0.10	0.23	0.31	0.66	9.26	0.04	0.05	0.03
s2_left_6_O14_G43N_L11+s2_right_5_O14_G43C_L11	0.12	0.09	0.13	0.23	0.84	11.45	0.07	0.03	0.03
s2_left_6_O14_G43N_L9+s2_right_5_O14_G43C_L11	0.06	0.06	0.12	0.19	0.88	12.06	0.03	0.03	0.03
s2_left_6_O14_G43N_L6+s2_right_5_O14_G43C_L11	0.05	0.10	0.16	0.21	0.70	10.69	0.06	0.07	0.01
s2_left_6_O14_G43N_L4+s2_right_5_O14_G43C_L11	0.07	0.10	0.09	0.21	0.50	10.01	0.06	0.03	0.02
s2_left_6_O14_G43N_L26+s2_right_5_O14_G43C_L9	0.67	0.28	0.10	0.25	1.07	10.30	0.05	0.04	0.02
s2_left_6_O14_G43N_L21+s2_right_5_O14_G43C_L9	0.45	0.13	0.10	0.18	0.84	9.86	0.02	0.06	0.02
s2_left_6_O14_G43N_L18+s2_right_5_O14_G43C_L9	0.44	0.15	0.08	0.27	0.96	11.52	0.04	0.05	0.03
s2_left_6_O14_G43N_L13+s2_right_5_O14_G43C_L9	0.18	0.12	0.13	0.18	0.49	9.08	0.03	0.03	0.02
s2_left_6_O14_G43N_L11+s2_right_5_O14_G43C_L9	0.14	0.11	0.12	0.17	0.73	11.60	0.03	0.05	0.02
s2_left_6_O14_G43N_L9+s2_right_5_O14_G43C_L9	0.03	0.07	0.07	0.11	0.71	11.38	0.02	0.04	0.02
s2_left_6_O14_G43N_L6+s2_right_5_O14_G43C_L9	0.05	0.06	0.11	0.15	0.45	9.53	0.03	0.05	0.03
s2_left_6_O14_G43N_L4+s2_right_5_O14_G43C_L9	0.05	0.09	0.09	0.11	0.37	9.77	0.03	0.06	0.02

FIG. 18 (cont.)

	%Base Editing								
	G2	G5	C6	C8	**G10**	**G11**	G14	C15	C16
s2_left_6_O14_G43N_L26+s2_right_5_O14_G43C_L6	0.68	0.50	0.09	0.28	1.33	11.71	0.01	0.07	0.02
s2_left_6_O14_G43N_L21+s2_right_5_O14_G43C_L6	0.58	0.34	0.11	0.27	1.23	12.26	0.02	0.07	0.01
s2_left_6_O14_G43N_L18+s2_right_5_O14_G43C_L6	0.48	0.26	0.17	0.30	1.35	12.36	0.06	0.07	0.02
s2_left_6_O14_G43N_L13+s2_right_5_O14_G43C_L6	0.18	0.12	0.11	0.15	0.77	10.52	0.04	0.04	0.03
s2_left_6_O14_G43N_L11+s2_right_5_O14_G43C_L6	0.14	0.14	0.11	0.26	1.02	13.23	0.05	0.06	0.02
s2_left_6_O14_G43N_L9+s2_right_5_O14_G43C_L6	0.08	0.14	0.15	0.19	1.04	14.14	0.03	0.05	0.04
s2_left_6_O14_G43N_L6+s2_right_5_O14_G43C_L6	0.03	0.16	0.09	0.17	0.73	12.31	0.03	0.05	0.03
s2_left_6_O14_G43N_L4+s2_right_5_O14_G43C_L6	0.04	0.10	0.07	0.15	0.48	10.91	0.03	0.05	0.02
s2_left_6_O14_G43N_L26+s2_right_5_O14_G43C_L4	0.88	1.00	0.14	0.32	1.89	15.16	0.02	0.04	0.03
s2_left_6_O14_G43N_L21+s2_right_5_O14_G43C_L4	0.60	0.47	0.17	0.38	1.68	15.03	0.04	0.06	0.02
s2_left_6_O14_G43N_L18+s2_right_5_O14_G43C_L4	0.61	0.33	0.13	0.38	1.72	14.64	0.03	0.06	0.05
s2_left_6_O14_G43N_L13+s2_right_5_O14_G43C_L4	0.21	0.23	0.10	0.21	1.15	13.10	0.03	0.05	0.03
s2_left_6_O14_G43N_L11+s2_right_5_O14_G43C_L4	0.14	0.23	0.13	0.21	1.28	15.04	0.03	0.05	0.02
s2_left_6_O14_G43N_L9+s2_right_5_O14_G43C_L4	0.08	0.17	0.14	0.20	1.35	15.17	0.02	0.03	0.02
s2_left_6_O14_G43N_L6+s2_right_5_O14_G43C_L4	0.06	0.17	0.15	0.20	0.89	14.01	0.02	0.04	0.02
s2_left_6_O14_G43N_L4+s2_right_5_O14_G43C_L4	0.06	0.17	0.13	0.23	0.61	13.35	0.03	0.06	0.03

FIG. 18 (cont.)

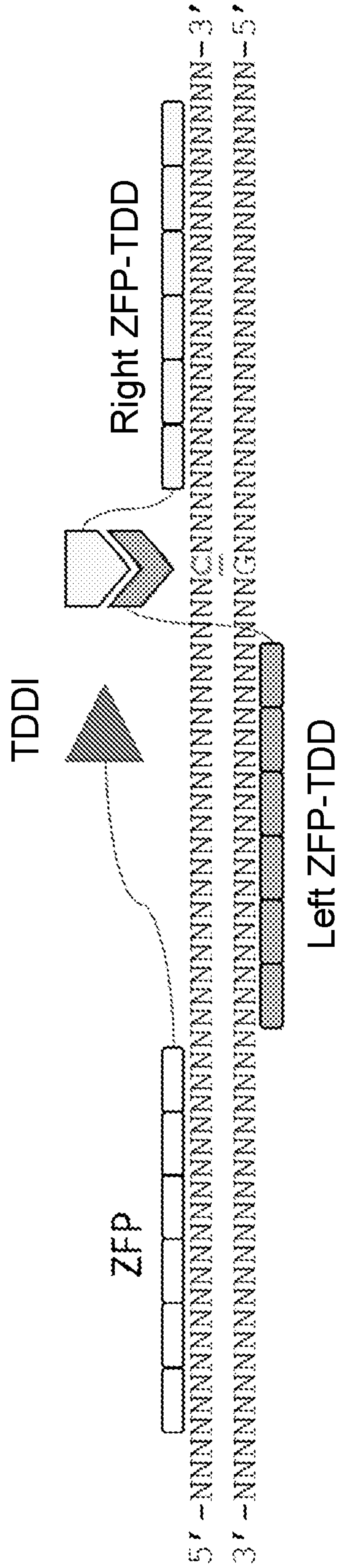


FIG. 19