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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
01 December 2022 (01.12.2022)

WIPO | PCT

(10) International Publication Number
WO 2022/251687 A3

(51) International Patent Classification:
A61K 31/70 (2006.01) *C12N 15/90* (2006.01)
C12N 15/10 (2006.01) *A61K 38/46* (2006.01)
C12N 15/11 (2006.01)

(21) International Application Number:
PCT/US2022/031419

(22) International Filing Date:
27 May 2022 (27.05.2022)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/194,431 28 May 2021 (28.05.2021) US

(71) Applicant: **BEAM THERAPEUTICS INC.** [US/US]; 238 Main Street, 9th Floor, Cambridge, Massachusetts 02142 (US).

(72) Inventors: **BRYSON, David**; c/o Beam Therapeutics Inc., 238 Main Street, 9th Floor, Cambridge, Massachusetts 02142 (US). **SULLIVAN, Jack**; c/o Beam Therapeutics Inc., 238 Main Street, 9th Floor, Cambridge, Massachusetts 02142 (US).

(74) Agent: **HUNTER-ENSOR, Melissa**; c/o Greenberg Traurig, LLP, One International Place, Boston, Massachusetts 02110 (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, IQ, IR, IS, IT, JM, 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))*
— *of inventorship (Rule 4.17(iv))*

Published:
— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*

(88) Date of publication of the international search report:
19 January 2023 (19.01.2023)

(54) Title: COMPOSITIONS AND METHODS FOR THE SELF-INACTIVATION OF BASE EDITORS

(57) Abstract: Disclosed herein are polynucleotides encoding a deaminase or napDNAbp polypeptide comprising an intron inserted in an open-reading frame encoding the deaminase or napDNAbp, further wherein the intron has an altered splice acceptor or splice donor site which functions to decrease splicing of editing mRNA. Also disclosed are polynucleotides encoding a base editor open reading frame comprising an intron, wherein the base editor comprises a napDNAbp domain or a deaminase domain

WO 2022/251687 A3

COMPOSITIONS AND METHODS FOR THE SELF-INACTIVATION OF BASE EDITORS

CROSS-REFERENCE TO RELATED APPLICATION

This application claims priority to and the benefit of U.S. Provisional Application No. 63/194,431, filed May 28, 2021, the entire contents of which is incorporated herein by reference.

SEQUENCE LISTING

This application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on May 27, 2022 is named 180802_049001_PCT_SL.txt and is 2,089,884 bytes in size.

BACKGROUND OF THE INVENTION

Advances in gene-editing technologies, such as the application of CRISPR-Cas systems in eukaryotes and the advent of base editing, allow the genome to be efficiently edited in a wide variety of cell types and organisms, rapidly expanding the available approaches to treat genetic diseases in humans. Although CRISPR-Cas systems and base editors can be highly specific for a genomic target of interest, transient expression of genome-modifying tools in cells is preferred in order to mitigate potential off-target editing events more likely to occur if expression were to persist over a longer period. Thus, methods to subsequently inhibit or halt editing activity after successful on-target editing are of broad interest particularly when delivery methods are utilized that may result in long-term expression, such as through adeno-associated virus (AAV) transduction, DNA transfection, or other methods.

SUMMARY OF THE INVENTION

As described below, the present invention features self-inactivating base editors and related compositions and methods.

In one aspect, the invention of the disclosure features a polynucleotide encoding a deaminase domain or a nucleic acid programmable DNA binding protein (napDNAbp) domain or fragment thereof. The polynucleotide contains an intron. The intron is inserted in an open reading frame encoding the deaminase or a napDNAbp or fragment thereof.

In another aspect, the invention of the disclosure features a polynucleotide encoding a deaminase domain or a nucleic acid programmable DNA binding protein (napDNAbp) domain open reading frame containing an intron. The intron contains an alteration at a splice acceptor or

splice donor site. The alteration reduces or eliminates splicing of base editor mRNA, thereby reducing or eliminating expression of a base editor polypeptide.

In another aspect, the invention of the disclosure features a polynucleotide encoding a base editor polypeptide or fragment thereof. The polynucleotide contains an intron. The intron is inserted in an open reading frame encoding the base editor polypeptide or fragment thereof.

In another aspect, the invention of the disclosure features a polynucleotide containing a base editor open reading frame containing an intron. The intron contains an alteration at a splice acceptor or splice donor site. The alteration reduces or eliminates splicing of base editor mRNA, thereby reducing or eliminating expression of a base editor polypeptide.

In another aspect, the invention of the disclosure features a polynucleotide encoding a base editor containing a nucleic acid programmable DNA binding protein (napDNAbp) domain or a deaminase domain. The polynucleotide contains an intron. The intron is inserted in an open reading frame encoding the napDNAbp domain or the deaminase domain.

In another aspect, the invention of the disclosure features a polynucleotide encoding a base editor containing a nucleic acid programmable DNA binding protein (napDNAbp) domain, and a deaminase domain, or a fragment thereof. The polynucleotide contains a base editor open reading frame containing an intron. The intron contains an alteration at a splice acceptor or splice donor site. The alteration reduces splicing of the base editor mRNA.

In another aspect, the invention of the disclosure features a composition containing (i) a first polynucleotide encoding a deaminase domain and an N- terminal fragment of a nucleic acid programmable DNA binding protein (napDNAbp) domain, where the N-terminal fragment of the napDNAbp domain is fused to a split intein-N. The composition also contains (ii) a second polynucleotide encoding a C-terminal fragment of the napDNAbp domain, where the C- terminal fragment of the napDNAbp domain is fused to a split intein-C. The first or second polynucleotide contains an intron, where the intron is inserted in an open reading frame of the polynucleotides.

In another aspect, the invention of the disclosure features a composition containing (i) a first polynucleotide encoding an N-terminal fragment of a deaminase domain, where the N-terminal fragment of the deaminase domain is fused to a split intein-N. The composition also contains (ii) a second polynucleotide encoding a C-terminal fragment of the deaminase domain and a nucleic acid programmable DNA binding protein (napDNAbp) domain, where the C-terminal fragment of the deaminase domain is fused to a split intein-C. The first or second polynucleotide contains an intron, where the intron is inserted in an open reading frame of the polynucleotides.

In another aspect, the invention of the disclosure features a base editor system containing (i) a polynucleotide encoding a base editor containing a deaminase domain, or fragment thereof. The base editor system also contains (ii) one or more guide RNAs that direct the base editor to edit a site in the genome of a cell. The base editor system further contains (iii) one or more guide RNAs that direct the base editor to edit the polynucleotide encoding the base editor. The edit results in a decrease in activity and/or expression of the encoded base editor.

In another aspect, the invention of the disclosure features a base editor system containing (i) a polynucleotide encoding a self-inactivating base editor or fragment thereof, where the polynucleotide contains an intron inserted in an open reading frame of the self-inactivating base editor or fragment thereof. The base editor system further contains (ii) one or more guide RNAs that direct the self-inactivating base editor to edit a site in the genome of a cell. The base editor system also contains (iii) one or more guide RNAs that direct the self-inactivating base editor to edit a splice acceptor or a splice donor site present in the intron of the polynucleotide encoding the self-inactivating base editor.

In another aspect, the invention of the disclosure features a base editor system containing (i) the polynucleotide of any one of the above aspects encoding a base editor. The base editor system also contains (ii) one or more guide RNAs that direct the base editor to edit a site in the genome of a cell. The base editor system further contains (iii) one or more guide RNAs that direct the base editor to edit a splice acceptor or a splice donor site present in the intron of the polynucleotide encoding the base editor.

In another aspect, the invention of the disclosure features a base editor system containing (i) the composition of any of the above aspects encoding a base editor. The base editor system further contains (ii) one or more guide RNAs that direct the base editor to edit a site in the genome of a cell. The base editor system also contains (iii) one or more guide RNAs that direct the base editor to edit a splice acceptor or a splice donor site present in the intron of the composition of (i).

In another aspect, the invention of the disclosure features a base editor system containing (i) a first polynucleotide encoding a deaminase domain and an N-terminal fragment of a nucleic acid programmable DNA binding protein (napDNAbp) domain, where the N-terminal fragment of the napDNAbp domain is fused to a split intein-N. The base editor system also contains (ii) a second polynucleotide encoding a C-terminal fragment of the napDNAbp domain, where the C-terminal fragment of the napDNAbp domain is fused to a split intein-C. The first or second polynucleotide contains an intron, where the intron is inserted in an open reading frame, and where the first and second polynucleotides encode a base editor. The base editor system further

contains (iii) one or more guide RNAs that direct the base editor to edit a site in the genome of a cell. The base editor system also contains (iv) one or more guide RNAs that direct the base editor to edit a splice acceptor or a splice donor site present in the intron of the polynucleotide of (i) or (ii).

In another aspect, the invention of the disclosure features a base editor system containing (i) a first polynucleotide encoding an N- terminal fragment of a deaminase domain, where the N-terminal fragment of the deaminase domain is fused to a split intein-N. The base editor system also contains (ii) a second polynucleotide encoding a C-terminal fragment of the deaminase domain and a nucleic acid programmable DNA binding protein (napDNAbp) domain, where the C- terminal fragment of the deaminase domain is fused to a split intein-C. The first or second polynucleotide contains an intron, where the intron is inserted in an open reading frame, and where the first and second polynucleotides encode a base editor. The base editor system also contains (iii) one or more guide RNAs that direct the base editor to edit a site in the genome of a cell. The base editor system also contains (iv) one or more guide RNAs that direct the base editor to edit a splice acceptor or a splice donor site present in the intron of the polynucleotide of (i) or (ii).

In another aspect, the invention of the disclosure features a vector containing a polynucleotide encoding a self-inactivating base editor or fragment thereof. The polynucleotide contains an intron inserted in an open reading frame of the self-inactivating base editor or fragment thereof.

In another aspect, the invention of the disclosure features a vector containing the polynucleotide of any of the above aspects, or embodiments thereof, or the base editor system of any of the above aspects, or embodiments thereof.

In another aspect, the invention of the disclosure features a vector containing the first polynucleotide and/or the second polynucleotide of the composition of any one of the above aspects.

In another aspect, the invention of the disclosure features a cell containing a vector containing a polynucleotide encoding a self-inactivating base editor or fragment thereof. The polynucleotide contains an intron inserted in an open reading frame of the self-inactivating base editor or fragment thereof.

In another aspect, the invention of the disclosure features a cell containing the polynucleotide of any of the above aspects, or embodiments thereof, the composition of any of the above aspects, or embodiments thereof, the base editor system of any of the above aspects, or embodiments thereof, or the vector of any of the above aspects, or embodiments thereof.

In another aspect, the invention of the disclosure features a pharmaceutical composition containing the polynucleotide of any of the above aspects, or embodiments thereof, the base editor system of any of the above aspects, or embodiments thereof, the vector of any of the above aspects, or embodiments thereof, or the cell of any of the above aspects, or embodiments thereof.

In another aspect, the invention of the disclosure features a kit containing the polynucleotide, the composition, the base editor system, the vector, the cell, or the pharmaceutical composition of any of the above aspects, or embodiments thereof.

In another aspect, the invention of the disclosure features a method for reducing or eliminating expression of a self-inactivating base editor. The method involves (a) providing a polynucleotide encoding a self-inactivating base editor or fragment thereof, where the polynucleotide contains an intron inserted in an open reading frame of the self-inactivating base editor or fragment thereof. The method also involves (b) contacting the polynucleotide with a guide RNA and a self-inactivating base editor polypeptide, where the guide RNA directs the base editor to edit a splice acceptor or a splice donor site of the intron, thereby generating an alteration that reduces or eliminates expression of the self-inactivating base editor.

In another aspect, the invention of the disclosure features a method of self-inactivating base editing. The method involves (a) expressing in a cell a polynucleotide encoding a base editor containing a deaminase domain, or fragment thereof. The method also involves (b) contacting the cell with a first guide RNA that directs the base editor to edit a site in the genome of the cell, thereby generating an alteration in the genome of the cell. The method further involves (c) contacting the cell with a second guide RNA that directs the base editor to edit the polynucleotide encoding the base editor, where the edit results in a decrease in activity and/or expression of the encoded base editor, thereby generating an alteration that reduces or eliminates expression of the base editor.

In another aspect, the invention of the disclosure features a method of self-inactivating base editing. The method involves (a) expressing in a cell a polynucleotide encoding a self-inactivating base editor or fragment thereof, where the polynucleotide contains an intron inserted in an open reading frame of the self-inactivating base editor or fragment thereof. The method also involves (b) contacting the cell with a first guide RNA that directs the self-inactivating base editor to edit a site in the genome of the cell, thereby generating an alteration in the genome of the cell. The method further involves (c) contacting the cell with a second guide RNA that directs the self-inactivating base editor to edit a splice acceptor or a splice donor site present in the intron of the polynucleotide of (a), thereby generating an alteration that reduces or eliminates expression of the self-inactivating base editor.

In another aspect, the invention of the disclosure features a method of editing the genome of an organism. The method involves (a) expressing in a cell of the organism a polynucleotide encoding a self-inactivating base editor or fragment thereof, where the polynucleotide contains an intron inserted in an open reading frame of the self-inactivating base editor or fragment thereof. The method also involves (b) contacting the cell with a first guide RNA that directs the self-inactivating base editor to edit a site in the genome of the cell, thereby generating an alteration in the genome of the cell. The method further involves (c) contacting the cell with a second guide RNA that directs the self-inactivating base editor to edit a splice acceptor or a splice donor site present in the intron of the polynucleotide of (a), thereby generating an alteration that reduces or eliminates expression of the self-inactivating base editor.

In another aspect, the invention of the disclosure features a method of treating a subject. The method involves (a) expressing in a cell of the subject a polynucleotide encoding a self-inactivating base editor or fragment thereof, where the polynucleotide contains an intron inserted in an open reading frame of the self-inactivating base editor or fragment thereof. The method further involves (b) contacting the cell with a first guide RNA that directs the self-inactivating base editor to edit a site in the genome of the cell, thereby generating an alteration in the genome of the cell to treat the subject. The method also involves (c) contacting the cell with a second guide RNA that directs the self-inactivating base editor to edit a splice acceptor or a splice donor site present in the intron of the polynucleotide of (a), thereby generating an alteration that reduces or eliminates expression of the self-inactivating base editor.

In another aspect, the invention of the disclosure features a method of treating a subject. The method involves administering to the subject the base editor system, the vector, the cell, or the pharmaceutical composition of any of the above aspects, or embodiments thereof, thereby treating the subject.

In another aspect, the invention of the disclosure features a method of editing the genome of an organism. The method involves (a) expressing in a cell of the organism a first polynucleotide encoding a deaminase domain and an N-terminal fragment of a nucleic acid programmable DNA binding protein (napDNAbp) domain, where the N-terminal fragment of the napDNAbp domain is fused to a split intein-N, and a second polynucleotide encoding a C-terminal fragment of the napDNAbp domain, where the C- terminal fragment of the napDNAbp domain is fused to a split intein-C. The first or second polynucleotide contains an intron. The intron is inserted in an open reading frame. Expression of the first and second polynucleotides in the cell result in the formation of a self-inactivating base editor. The method also involves (b) contacting the cell with a first guide RNA that directs the self-inactivating base editor to edit a

site in the genome of the cell, thereby generating an alteration in the genome of the cell. The method also involves (c) contacting the cell with a second guide RNA that directs the self-inactivating base editor to edit a splice acceptor or a splice donor site present in the intron of the polynucleotide of (a), thereby generating an alteration that reduces or eliminates expression of the self-inactivating base editor.

In another aspect, the invention of the disclosure features a method of editing the genome of an organism. The method involves (a) expressing in a cell of the organism a first polynucleotide encoding an N-terminal fragment of a deaminase domain, where the N-terminal fragment of the deaminase domain is fused to a split intein-N, and a second polynucleotide encoding a C-terminal fragment of the deaminase domain and a nucleic acid programmable DNA binding protein (napDNAbp) domain, where the C- terminal fragment of the deaminase domain is fused to a split intein-C. The first or second polynucleotide contains an intron, where the intron is inserted in an open reading frame. Expression of the first and second polynucleotides in the cell result in the formation of a self-inactivating base editor. The method also involves (b) contacting the cell with a first guide RNA that directs the self-inactivating base editor to edit a site in the genome of the cell, thereby generating an alteration in the genome of the cell. The method further involves (c) contacting the cell with a second guide RNA that directs the self-inactivating base editor to edit a splice acceptor or a splice donor site present in the intron of the polynucleotide of (a), thereby generating an alteration that reduces or eliminates expression of the self-inactivating base editor.

In any of the above aspects, or embodiments thereof, the base editor has high editing efficiency in genomic DNA. In any of the above aspects, or embodiments thereof, the base editor contains a nucleic acid programmable DNA binding protein (napDNAbp) domain or a deaminase domain.

In any of the above aspects, or embodiments thereof, the deaminase domain is a cytidine deaminase domain or an adenosine deaminase domain. In any of the above aspects, or embodiments thereof, the deaminase domain is a TadA domain.

In any of the above aspects, or embodiments thereof, the napDNAbp domain is a Cas domain selected from one or more of a Cas9, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, and Cas12j/CasΦ domain.

In any of the above aspects, or embodiments thereof, the intron is derived from a sequence selected from one or more of NF1, PAX2, EEF1A1, HBB, IGHG1, SLC50A1, ABCB11, BRSK2, PLXNB3, TMPRSS6, IL32, ANTXRL, PKHD1L1, PADI1, KRT6C, and HMCN2. In any of the above aspects, or embodiments thereof, the intron is derived from NF1.

In any of the above aspects, or embodiments thereof, the intron is derived from PAX2. In any of the above aspects, or embodiments thereof, the intron is derived from EEF1A1. In any of the above aspects, or embodiments thereof, the intron is derived from HBB. In any of the above aspects, or embodiments thereof, the intron is derived from IGHG1. In any of the above aspects, or embodiments thereof, the intron is derived from SLC50A1. In any of the above aspects, or embodiments thereof, the intron is derived from ABCB11. In any of the above aspects, or embodiments thereof, the intron is derived from BRSK2. In any of the above aspects, or embodiments thereof, the intron is derived from PLXNB3. In any of the above aspects, or embodiments thereof, the intron is derived from TMPRSS6. In any of the above aspects, or embodiments thereof, the intron is derived from IL32. In any of the above aspects, or embodiments thereof, the intron is derived from PKHD1L1. In any of the above aspects, or embodiments thereof, the intron is derived from PADI1. In any of the above aspects, or embodiments thereof, the intron is derived from KRT6C. In any of the above aspects, or embodiments thereof, the intron is derived from HMCN2. In any of the above aspects, or embodiments thereof, the intron has at least about 85% nucleic acid sequence identity to an intron naturally present in a mammalian gene. In any of the above aspects, or embodiments thereof, the intron has at least about 85% nucleic acid sequence identity to an intron naturally present in a non-mammalian gene. In any of the above aspects, or embodiments thereof, the intron is a synthetic intron. In any of the above aspects, or embodiments thereof, the intron contains a sequence that has at least about 85% nucleic acid sequence identity to one of the following:

a) GTGAGATCAAATGAAAGTTTCATATAGAAATACAAAACCTAGAGAACTGGCATGTAAG AGAAGCAAAAATTACTTCAGCAAGGCCATGTTAGTAAATTTGCATCTGTTTGTCCACATTAG (SEQ ID NO: 226);

b) GTAGGTGACAATGCTGCAGCTGCCTAATCTAGGTGGGGGGAACTAAATTGTGGGTGAG CTGCTGAATGGTCTGTAGTCTGAGGCTGGGGTGGGGGGAGACACAACGTCCCCTCCCTGCAAAC CACTGCTATTCTGTCCCTCTCTCTCCTTAG (SEQ ID NO: 227);

c) GTAAGTGGCTTTCAAGACCATTGTTAAAAAGCTCTGGGAATGGCGATTTCATGCTTAC ATAAATTGGCATGCTTGTGTTTCAG (SEQ ID NO: 228);

d) GTAAGTATCAAGGTTACAAGACAGGTTTAAGGAGACCAATAGAAACTGGGCTTGTCTA GACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTT TCTCTCCACAG (SEQ ID NO: 229);

e) GTAAGCACAACTGGGATGGGGTGACAGGGGTGCAAGATTGAAAACTGGCTCCTCTCCT CATAGCAGTTCTTGTGATTTCAG (SEQ ID NO: 230);

f) GTAAGAAATGTTATTTTTCAGTAAGTGATTTAGTTATTTTTCCTTTTTTTCTCATTAAA ATTTCTCTAACATCTCCCTCTTCATGTTTTAG (SEQ ID NO: 231);

g) GTGAGACCCTAGCCCCCTCAACCCTGCCCTGGCCTCTCCCCAAACCTGCCCCCCCACG CTGACCCCCACACCCGGCCGCCCGCAG (SEQ ID NO: 232);

h) GTGGGTGTCAGAGGCATCGGGGCTGCGGGGTAGGGGGCTGCCCCACCCCTAACGAAGT CTGCTCCTCCAG (SEQ ID NO: 233);

i) GCAGGGAAGTCCTGCTTCCGTGCCCCACCGGTGCTCAGCTGAGGCTCCCTTGAAAATG CGAGGCTGTTTCCAACTTTGGTCTGTTTCCCTGGCAG (SEQ ID NO: 234);

j) GTGGGGAGTTGGGGTCCCCGAAGGTGAGGACCCTCTGGGGATGAGGGTGCTTCTCTGA GACACTTTCTTTTCCTCACACCTGTTCCTCGCCAGCAG (SEQ ID NO: 235);

k) GTATAGACCCCTTGATCTCCTAACCCTAACCCTAACCCTAACCCTAACCTACAAAATC TTAGAGCATCAGTGGGAGCATCTCACTGTCCAGGCTCAATATTTCTTCATTTTCTTGCAG (SEQ ID NO: 236);

l) GTAATTATGATTAAAGATGGTGATTGTTTATTTTCTTTTATGATTGTCCTTAGTATTA TGTAACCTGCAAATTCTATTGCAG (SEQ ID NO: 237);

m) GTGAGTGACACAAGGTGTTGTCTGGGGAGTGGGGAAGGGGGATGGAAGTGAATCCTG TTGGTGGGGTGGAGAAAGGGCGATCTCAAGAGGGCCACTCTCTCCAG (SEQ ID NO: 238);

n) GTAAGCATCTCCACCATCCTTCTGTTTACTCTGATGGGGTCTGCAAAGGGGAGATGAT GTATAGGGTTGGGTATCTCTGTAAATGTCAGATGTGAAGTTGATCTTATGACCTTCTGTTCTGC AG (SEQ ID NO: 239);

o) GTGAGGGTCTCCCAGGCTGGGCAGGGGGAGGGGGCTGCTGCCTTGATTGCGTCCCAGG ACACAGCCCTCCTCCAGCCTGCCCTCGCCTTGCTCATCCCCTCCCCATCTCAGCCCCACCCCCA CTAACTCTCTCTCTGCTCTGACTCAG (SEQ ID NO: 240);

p) GTAATGATTGATTGCAATGTATGATTACAATAATCTCAGTATAAGTTCAGTAATAATA ACCTTCCACTGCTGTCCTCTGTGTGCACCCAG (SEQ ID NO: 241); or

q) GTAAATATATACAACAGTTTTTCATTTAAATAAGTGCACGGCACAAATAAGAAAAAT ATGTCAAAAATGTAACCAATAGTTTTTTTCAAATTTAG (SEQ ID NO: 242).

In any of the above aspects, or embodiments thereof, the intron contains a nucleic acid sequence from one of the following:

a) GTGAGATCAAATGAAAGTTTCATATAGAAATACAAAACCTAGAGAACTGGCATGTAAG AGAAGCAAAAATTACTTCAGCAAGGCCATGTTAGTAAATTTGCATCTGTTTGTCCACATTAG (SEQ ID NO: 226);

b) GTAGGTGACAATGCTGCAGCTGCCTAATCTAGGTGGGGGGAACTAAATTGTGGGTGAG CTGCTGAATGGTCTGTAGTCTGAGGCTGGGGTGGGGGGAGACACAACGTCCCCTCCCTGCAAAC CACTGCTATTCTGTCCCTCTCTCTCCTTAG (SEQ ID NO: 227);

c) GTAAGTGGCTTTCAAGACCATTGTTAAAAAGCTCTGGGAATGGCGATTTCATGCTTAC ATAAATTGGCATGCTTGTGTTTCAG (SEQ ID NO: 228);

d) GTAAGTATCAAGGTTACAAGACAGGTTTAAGGAGACCAATAGAAACTGGGCTTGTCTA GACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTT TCTCTCCACAG (SEQ ID NO: 229);

e) GTAAGCACAACTGGGATGGGGTGACAGGGGTGCAAGATTGAAAACTGGCTCCTCTCCT CATAGCAGTTCTTGTGATTTCAG (SEQ ID NO: 230);

f) GTAAGAAATGTTATTTTTCAGTAAGTGATTTAGTTATTTTTCCTTTTTTCTCATTAAA ATTTCTCTAACATCTCCCTCTTCATGTTTTAG (SEQ ID NO: 231);

g) GTGAGACCCTAGCCCCCTCAACCCTGCCCTGGCCTCTCCCCAAACCTGCCCCCCCACG CTGACCCCCACACCCGGCCGCCCGCAG (SEQ ID NO: 232);

h) GTGGGTGTCAGAGGCATCGGGGCTGCGGGGTAGGGGGCTGCCCCACCCCTAACGAAGT CTGCTCCTCCAG (SEQ ID NO: 233);

i) GCAGGGAAGTCCTGCTTCCGTGCCCCACCGGTGCTCAGCTGAGGCTCCCTTGAAAATG CGAGGCTGTTTCCAACTTTGGTCTGTTTCCCTGGCAG (SEQ ID NO: 234);

j) GTGGGGAGTTGGGGTCCCCGAAGGTGAGGACCCTCTGGGGATGAGGGTGCTTCTCTGA GACACTTTCTTTTCCTCACACCTGTTCCTCGCCAGCAG (SEQ ID NO: 235);

k) GTATAGACCCCTTGATCTCCTAACCCTAACCCTAACCCTAACCCTAACCTACAAAATC TTAGAGCATCAGTGGGAGCATCTCACTGTCCAGGCTCAATATTTCTTCATTTTCTTGCAG (SEQ ID NO: 236);

l) GTAATTATGATTAAAGATGGTGATTGTTTATTTTCTTTTATGATTGTCCTTAGTATTA TGTAACCTGCAAATTCTATTGCAG (SEQ ID NO: 237);

m) GTGAGTGACACAAGGTGTTGTCTGGGGAGTGGGGAAGGGGGATGGAAGTGAATCCTG TTGGTGGGGTGGAGAAAGGGCGATCTCAAGAGGGCCACTCTCTCCAG (SEQ ID NO: 238);

n) GTAAGCATCTCCACCATCCTTCTGTTTACTCTGATGGGGTCTGCAAAGGGGAGATGAT GTATAGGGTTGGGTATCTCTGTAAATGTCAGATGTGAAGTTGATCTTATGACCTTCTGTTCTGC AG (SEQ ID NO: 239);

o) GTGAGGGTCTCCCAGGCTGGGCAGGGGGAGGGGGCTGCTGCCTTGATTGCGTCCCAGG ACACAGCCCTCCTCCAGCCTGCCCTCGCCTTGCTCATCCCCTCCCCATCTCAGCCCCACCCCCA CTAACTCTCTCTCTGCTCTGACTCAG (SEQ ID NO: 240);

p) GTAATGATTGATTGCAATGTATGATTACAATAATCTCAGTATAAGTTCAGTAATAATA ACCTTCCACTGCTGTCCTCTGTGTGCACCCAG (SEQ ID NO: 241); or

q) GTAAATATATACAACAGTTTTTCATTTAAATAAGTGCACGGCACAAATAAGAAAAAT ATGTCAAAAATGTAACCAATAGTTTTTTTCAAATTTAG (SEQ ID NO: 242).

In any of the above aspects, or embodiments thereof, the intron contains between about 10 base pairs to about 500 base pairs. In any of the above aspects, or embodiments thereof, the intron contains between about 70 base pairs and 150 base pairs. In any of the above aspects, or embodiments thereof, the intron contains between about 100 base pairs and 200 base pairs. In any of the above aspects, or embodiments thereof, the intron is inserted in proximity to a protospacer sequence. In any of the above aspects, or embodiments thereof, the intron is inserted within about 10 to 30 base pairs of the protospacer sequence. In any of the above aspects, or embodiments thereof, the protospacer sequence is NGG or NNGRRT.

In any of the above aspects, or embodiments thereof, the deaminase domain contains a TadA domain.

In any of the above aspects, or embodiments thereof, the intron is inserted within or directly after codon 18, 23, 59, 62, 87, or 129 of TadA. In any of the above aspects, or embodiments thereof, the intron is inserted directly after codon 87 of TadA. In any of the above aspects, or embodiments thereof, the alteration is a single-base edit. In any of the above aspects, or embodiments thereof, the single-base edit is an A-to-G base edit. In any of the above aspects, or embodiments thereof, the single-base edit is a C-to-T base edit.

In any of the above aspects, or embodiments thereof, the polynucleotide further contains a polynucleotide sequence encoding a linker. In any of the above aspects, or embodiments thereof, the intron is inserted within the polynucleotide sequence encoding the linker.

In any of the above aspects, or embodiments thereof, the programmable DNA binding protein domain is a Cas9 domain. In any of the above aspects, or embodiments thereof, the Cas9 domain is split between amino acid residues corresponding to Asn309 and Thr310 of Cas9, and residue 310 was mutated to a Thr310Cys.

In any of the above aspects, or embodiments thereof, the intron contains an alteration at a splice acceptor or splice donor site, where the alteration reduces or eliminates splicing of base editor mRNA.

In any of the above aspects, or embodiments thereof, the napDNAbp domain is a Cas9 domain. In any of the above aspects, or embodiments thereof, the N- and C-terminal domains of the Cas9 domain are split between amino acid residues Asn309 and Thr310. In any of the above aspects, or embodiments thereof, the Cas9 domain contains the mutation Thr310Cys.

In any of the above aspects, or embodiments thereof, the composition further contains a linker polynucleotide sequence. In any of the above aspects, or embodiments thereof, the intron is inserted within the linker polynucleotide sequence.

In any of the above aspects, or embodiments thereof, the edit alters a catalytic residue of the deaminase domain. In any of the above aspects, or embodiments thereof, the deaminase domain is an adenosine deaminase domain. In any of the above aspects, or embodiments thereof, the deaminase domain is and cytidine deaminase domain. In any of the above aspects, or embodiments thereof, the altered catalytic residue of the deaminase domain is His57 (H57), Glu59 (E59), Cys87 (C87), or Cys90 (C90) of the following reference sequence: MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH RVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 1), or a corresponding position in another adenosine deaminase. In any of the above aspects, or embodiments thereof, the altered catalytic residue is E59. In any of the above aspects, or embodiments thereof, the alteration to the catalytic residue is E59G. In any of the above aspects, or embodiments thereof, the altered catalytic residue is H57. In any of the above aspects, or embodiments thereof, the alteration to the catalytic residue is H57R. In any of the above aspects, or embodiments thereof, the altered catalytic residue is C87. In any of the above aspects, or embodiments thereof, the alteration to the catalytic residue is C87R. In any of the above aspects, or embodiments thereof, the altered catalytic residue is C90. In any of the above aspects, or embodiments thereof, the alteration to the catalytic residue is C90R.

In any of the above aspects, or embodiments thereof, the base editor system contains a polynucleotide sequence selected from the following:

a) gGUUUUAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 191);

b) gUUUCUUACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 192);

c) gGUUUCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 193);

d) GCCACUUACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 194);

e) gACAUUAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 195);

f) gGAUCUCACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 196);

g) gUCCUUAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 197);

h) GUCACCUACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 198);

i) GAUUUCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 190);

j) gGUGCUUACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 200);

k) gUCCACAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 201);

l) GAUACUUACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 202);

m) gUGUUUUAGCUGCGGCAAGGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 203);

n) gUUUCUUACAGCCAUAAUUUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 204);

o) gCUCCACAGCUGCGGCAAGGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 205);

p) GAUACUUACAGCCAUAAUUUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 206);

q) gUGUUUUAGGGACGAAAGAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 207);

r) gUUACCUGGCUCUCUUAGCCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 208);

s) gCUCCACAGGGACGAAAGAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 209);

t) gCUUGCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 210);

u) gAUUGCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 211);

v) gUCUCCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 212);

w) gUCUGCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 213);

x) gGACUCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 214);

y) GCACCCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 215);

z) gAAUUUAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 216);

aa) gCAUUAGGUCGAGAUCACAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 217);

bb) gCCUUAGGUCGAGAUCACAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 218);

cc) GUUUCAGGUCGAGAUCACAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 219);

dd) gACAUUAGGCUAAGAGAGCCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 220);

ee) gUCCUUAGGCUAAGAGAGCCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 221);

ff) gGUUUCAGGCUAAGAGAGCCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 222);

gg) gACAUUAGAUUAUGGCUCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 223);

hh) gUCCUUAGAUUAUGGCUCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 224);

ii) gGUUUCAGAUUAUGGCUCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 225);

jj) gCACCAUGAGCGAGGUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGU CCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 524);

kk) gGCCACCAUGAGCGAGGUCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 525);

ll) GUGUCGAAGUUCGCCCUGGAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 526);

mm) gAUGCCGAGAUAAUGGCCCUCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 527);

nn) gAUGCCGAGAUAAUGGCCCUUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 528);

oo) gAUGCCGAGAUCAUGGCACUAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 529);

pp) gAUGCCGAGAUCAUGGCACUCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 530);

qq) gAUGCCGAGAUCAUGGCACUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 531);

rr) gAUGCCGAGAUCAUGGCGCUAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 532);

ss) gAUGCCGAGAUCAUGGCGCUCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 533);

tt) gAUGCCGAGAUCAUGGCGUUAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 534);

uu) gAUGCCGAGAUUAUGGCACUAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 535);

vv) gAUGCCGAGAUUAUGGCACUCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 536);

ww) gAUGCCGAGAUUAUGGCACUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 537);

xx) gAUGCCGAGAUUAUGGCACUUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 538);

yy) gAUGCCGAGAUUAUGGCGCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 539);

zz) gAUGCCGAGAUUAUGGCUCUAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 540);

aaa) gAUGCGGAGAUCAUGGCGCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 541);

bbb) gAUGCUGAGAUAAUGGCCCUCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 542);

ccc) gAACCGCACAUGCCGAAAUUAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 543);

ddd) gGCAGGUGUCGACAUAUCUAUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 544);

eee) gAUGCCGAAAUUAUGGCUCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 545);

fff) gACACAUGACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 546); or

ggg) gGCCCCAGCACACAUGACACAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 547).

In any of the above aspects, or embodiments thereof, the expression vector is a mammalian expression vector. In any of the above aspects, or embodiments thereof, the vector is a lipid nanoparticle. In any of the above aspects, or embodiments thereof, the vector is a viral vector selected from one or more of an adeno-associated virus (AAV), retroviral vector, adenoviral vector, lentiviral vector, Sendai virus vector, and herpes virus vector. In any of the above aspects, or embodiments thereof, the vector is an AAV vector. In any of the above aspects, or embodiments thereof, the AAV vector is AAV2 or AAV8. In any of the above aspects, or embodiments thereof, the vector contains a promoter. In any of the above aspects, or embodiments thereof, the promoter is a CMV promoter.

In any of the above aspects, or embodiments thereof, the cell is *in vitro* or *in vivo*.

In any of the above aspects, or embodiments thereof, the composition or pharmaceutical composition further contains a pharmaceutically acceptable excipient, diluent, or carrier.

In any of the above aspects, or embodiments thereof, the kit contains instructions for use in the method of any of the above aspects, or embodiments thereof.

In any of the above aspects, or embodiments thereof, the method is performed *in vivo*.

In any of the above aspects, or embodiments thereof, the first polynucleotide and/or second polynucleotide are expressed in a cell by a vector. In any of the above aspects, or embodiments thereof, the first polynucleotide and second polynucleotide are expressed in a cell by separate vectors. In any of the above aspects, or embodiments thereof, the first guide RNA and/or second guide RNA are delivered to the cell by a vector. In any of the above aspects, or embodiments thereof, the first guide RNA and/or second guide RNA are delivered to the cell in the same vector than the first polynucleotide and/or second polynucleotide. In any of the above aspects, or embodiments thereof, the first guide RNA and/or second guide RNA are delivered to the cell in a different vector than the first polynucleotide and/or second polynucleotide. In any of the above aspects, or embodiments thereof, the vector is a viral vector.

In any of the above aspects, or embodiments thereof, the base editor contains a nucleic acid programmable DNA binding protein (napDNAbp) domain and a deaminase domain. In any of the above aspects, or embodiments thereof, the open reading frame containing the intron is in the napDNAbp domain or the deaminase domain.

In any of the above aspects, or embodiments thereof, the self-inactivating base editor polypeptide maintains high editing efficiency in genomic DNA. In any of the above aspects, or embodiments thereof, the deaminase domain is a cytidine deaminase domain or an adenosine deaminase domain. In any of the above aspects, or embodiments thereof, the alteration is in a consensus splice donor site at the 5' end of the intron or in a consensus splice acceptor sequence at the 3' end of the intron.

In any of the above aspects, or embodiments thereof, the intron contains a sequence that has at least about 85%, 90%, 95%, or 99% nucleic acid sequence identity to one of the following:

a) GTGAGATCAAATGAAAGTTTCATATAGAAATACAAAACCTAGAGAACTGGCATGTAAG AGAAGCAAAAATTACTTCAGCAAGGCCATGTTAGTAAATTTGCATCTGTTTGTCCACATTAG (SEQ ID NO: 226);

b) GTAGGTGACAATGCTGCAGCTGCCTAATCTAGGTGGGGGGAACTAAATTGTGGGTGAG CTGCTGAATGGTCTGTAGTCTGAGGCTGGGGTGGGGGGAGACACAACGTCCCCTCCCTGCAAAC CACTGCTATTCTGTCCCTCTCTCTCCTTAG (SEQ ID NO: 227);

c) GTAAGTGGCTTTCAAGACCATTGTTAAAAAGCTCTGGGAATGGCGATTTCATGCTTAC ATAAATTGGCATGCTTGTGTTTCAG (SEQ ID NO: 228);

d) GTAAGTATCAAGGTTACAAGACAGGTTTAAGGAGACCAATAGAAACTGGGCTTGTCTA GACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTT TCTCTCCACAG (SEQ ID NO: 229);

e) GTAAGCACAACTGGGATGGGGTGACAGGGGTGCAAGATTGAAAACTGGCTCCTCTCCT CATAGCAGTTCTTGTGATTTCAG (SEQ ID NO: 230);

f) GTAAGAAATGTTATTTTTCAGTAAGTGATTTAGTTATTTTTCCTTTTTTTCTCATTAAA ATTTCTCTAACATCTCCCTCTTCATGTTTTAG (SEQ ID NO: 231);

g) GTGAGACCCTAGCCCCCTCAACCCTGCCCTGGCCTCTCCCCAAACCTGCCCCCCCACG CTGACCCCCACACCCGGCCGCCCGCAG (SEQ ID NO: 232);

h) GTGGGTGTCAGAGGCATCGGGGCTGCGGGGTAGGGGGCTGCCCCACCCCTAACGAAGT CTGCTCCTCCAG (SEQ ID NO: 233);

i) GCAGGGAAGTCCTGCTTCCGTGCCCCACCGGTGCTCAGCTGAGGCTCCCTTGAAAATG CGAGGCTGTTTCCAACTTTGGTCTGTTTCCCTGGCAG (SEQ ID NO: 234);

j) GTGGGGAGTTGGGGTCCCCGAAGGTGAGGACCCTCTGGGGATGAGGGTGCTTCTCTGA GACACTTTCTTTTCCTCACACCTGTTCCTCGCCAGCAG (SEQ ID NO: 235);

k) GTATAGACCCCTTGATCTCCTAACCCTAACCCTAACCCTAACCCTAACCTACAAAATC TTAGAGCATCAGTGGGAGCATCTCACTGTCCAGGCTCAATATTTCTTCATTTTCTTGCAG (SEQ ID NO: 236);

l) GTAATTATGATTAAAGATGGTGATTGTTTATTTTCTTTTATGATTGTCCTTAGTATTA TGTAACCTGCAAATTCTATTGCAG (SEQ ID NO: 237);

m) GTGAGTGACACAAGGTGTTGTCTGGGGAGTGGGGAAGGGGGATGGAAGTGAATCCTG TTGGTGGGGTGGAGAAAGGGCGATCTCAAGAGGGCCACTCTCTCCAG (SEQ ID NO: 238);

n) GTAAGCATCTCCACCATCCTTCTGTTTACTCTGATGGGGTCTGCAAAGGGGAGATGAT GTATAGGGTTGGGTATCTCTGTAAATGTCAGATGTGAAGTTGATCTTATGACCTTCTGTTCTGC AG (SEQ ID NO: 239);

o) GTGAGGGTCTCCCAGGCTGGGCAGGGGGAGGGGGCTGCTGCCTTGATTGCGTCCCAGG ACACAGCCCTCCTCCAGCCTGCCCTCGCCTTGCTCATCCCCTCCCCATCTCAGCCCCACCCCCA CTAACTCTCTCTCTGCTCTGACTCAG (SEQ ID NO: 240);

p) GTAATGATTGATTGCAATGTATGATTACAATAATCTCAGTATAAGTTCAGTAATAATA ACCTTCCACTGCTGTCCTCTGTGTGCACCCAG (SEQ ID NO: 241); or

q) GTAAATATATACAACAGTTTTTCATTTAAATAAGTGCACGGCACAAATAAGAAAAAT ATGTCAAAAATGTAACCAATAGTTTTTTTCAAATTTAG (SEQ ID NO: 242).

In any of the above aspects, or embodiments thereof, the second guide RNA contains a polynucleotide sequence selected from the following:

a) gGUUUUAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 191);

b) gUUUCUUACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 192);

c) gGUUUCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 193);

d) GCCACUUACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 194);

e) gACAUUAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 195);

f) gGAUCUCACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 196);

g) gUCCUUAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 197);

h) GUCACCUACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 198);

i) GAUUUCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 190);

j) gGUGCUUACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 200);

k) gUCCACAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 201);

l) GAUACUUACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 202);

m) gUGUUUUAGCUGCGGCAAGGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 203);

n) gUUUCUUACAGCCAUAAUUUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 204);

o) gCUCCACAGCUGCGGCAAGGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 205);

p) GAUACUUACAGCCAUAAUUUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 206);

q) gUGUUUUAGGGACGAAAGAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 207);

r) gUUACCUGGCUCUCUUAGCCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 208);

s) gCUCCACAGGGACGAAAGAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 209);

t) gCUUGCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 210);

u) gAUUGCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 211);

v) gUCUCCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 212);

w) gUCUGCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 213);

x) gGACUCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 214);

y) GCACCCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 215);

z) gAAUUUAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 216);

aa) gCAUUAGGUCGAGAUCACAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 217);

bb) gCCUUAGGUCGAGAUCACAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 218);

cc) GUUUCAGGUCGAGAUCACAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 219);

dd) gACAUUAGGCUAAGAGAGCCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 220);

ee) gUCCUUAGGCUAAGAGAGCCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 221);

ff) gGUUUCAGGCUAAGAGAGCCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 222);

gg) gACAUUAGAUUAUGGCUCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 223);

hh) gUCCUUAGAUUAUGGCUCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 224);

ii) gGUUUCAGAUUAUGGCUCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 225);

jj) gCACCAUGAGCGAGGUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGU CCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 524);

kk) gGCCACCAUGAGCGAGGUCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 525);

ll) GUGUCGAAGUUCGCCCUGGAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 526);

mm) gAUGCCGAGAUAAUGGCCCUCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 527);

nn) gAUGCCGAGAUAAUGGCCCUUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 528);

oo) gAUGCCGAGAUCAUGGCACUAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 529);

pp) gAUGCCGAGAUCAUGGCACUCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 530);

qq) gAUGCCGAGAUCAUGGCACUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 531);

rr) gAUGCCGAGAUCAUGGCGCUAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 532);

ss) gAUGCCGAGAUCAUGGCGCUCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 533);

tt) gAUGCCGAGAUCAUGGCGUUAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 534);

uu) gAUGCCGAGAUUAUGGCACUAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 535);

vv) gAUGCCGAGAUUAUGGCACUCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 536);

ww) gAUGCCGAGAUUAUGGCACUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 537);

xx) gAUGCCGAGAUUAUGGCACUUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 538);

yy) gAUGCCGAGAUUAUGGCGCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 539);

zz) gAUGCCGAGAUUAUGGCUCUAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 540);

aaa) gAUGCGGAGAUCAUGGCGCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 541);

bbb) gAUGCUGAGAUAAUGGCCCUCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 542);

ccc) gAACCGCACAUGCCGAAAUUAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 543);

ddd) gGCAGGUGUCGACAUAUCUAUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 544);

eee) gAUGCCGAAAUUAUGGCUCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 545);

fff) gACACAUGACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 546); or

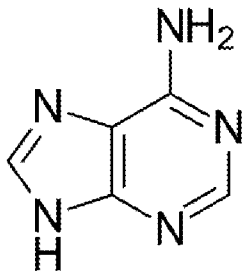
ggg) gGCCCCAGCACACAUGACACAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 547).

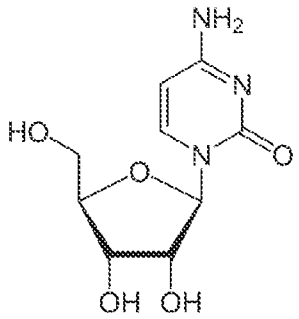
In any of the above aspects, or embodiments thereof, the polynucleotide further contains a linker polynucleotide sequence. In any of the above aspects, or embodiments thereof, the intron is inserted within the linker polynucleotide sequence.

In any of the above aspects, or embodiments thereof, the subject or organism is human. In any of the above aspects, or embodiments thereof, the subject or organism is a mammal. In any of the above aspects, or embodiments thereof, the mammal is human.

Definitions

Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by a person skilled in the art to which this invention belongs. The following references provide one of skill with a general definition of many of the terms used in this invention: Singleton *et al.*, Dictionary of Microbiology and Molecular Biology (2nd ed. 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988); The Glossary of Genetics, 5th Ed., R. Rieger *et al.* (eds.), Springer Verlag (1991); and Hale & Marham, The Harper Collins Dictionary of Biology (1991). As used herein, the following terms have the meanings ascribed to them below, unless specified otherwise.

By "adenine" or " 9*H*-Purin-6-amine" is meant a purine nucleobase with the molecular formula $C_5H_5N_5$, having the structure , and corresponding to CAS No. 73-24-5.

By "adenosine" or " 4-Amino-1-[(2*R*,3*R*,4*S*,5*R*)-3,4-dihydroxy-5-(hydroxymethyl)oxolan-2-yl]pyrimidin-2(1*H*)-one" is meant an adenine molecule attached to a ribose sugar via a glycosidic bond, having the structure , and corresponding to CAS No. 65-46-3. Its molecular formula is $C_{10}H_{13}N_5O_4$.

By "adenosine deaminase" or "adenine deaminase" is meant a polypeptide or functional fragment thereof capable of catalyzing the hydrolytic deamination of adenine or adenosine. The terms "adenine deaminase" and "adenosine deaminase" are used interchangeably throughout the application. In some embodiments, the deaminase or deaminase domain is an adenosine deaminase catalyzing the hydrolytic deamination of adenosine to inosine or deoxy adenosine to deoxyinosine. In some embodiments, the adenosine deaminase catalyzes the hydrolytic deamination of adenine or adenosine in deoxyribonucleic acid (DNA). The adenosine

deaminases (*e.g.* engineered adenosine deaminases, evolved adenosine deaminases) provided herein may be from any organism (*e.g.*, eukaryotic, prokaryotic), including but not limited to algae, bacteria, fungi, plants, invertebrates (*e.g.*, insects), and vertebrates (*e.g.*, amphibians, mammals). In some embodiments, the adenosine deaminase is an adenosine deaminase variant with one or more alterations and is capable of deaminating both adenine and cytosine in a target polynucleotide (*e.g.*, DNA, RNA). In some embodiments, the target polynucleotide is single or double stranded. In some embodiments, the adenosine deaminase variant is capable of deaminating both adenine and cytosine in DNA. In some embodiments, the adenosine deaminase variant is capable of deaminating both adenine and cytosine in single-stranded DNA. In some embodiments, the adenosine deaminase variant is capable of deaminating both adenine and cytosine in RNA.

By "adenosine deaminase activity" is meant catalyzing the deamination of adenine or adenosine to guanine in a polynucleotide. In some embodiments, an adenosine deaminase variant as provided herein maintains adenosine deaminase activity (*e.g.*, at least about 30%, 40%, 50%, 60%, 70%, 80%, 90% or more of the activity of a reference adenosine deaminase (*e.g.*, TadA*8.20 or TadA*8.19)).

By "Adenosine Base Editor (ABE)" is meant a base editor comprising an adenosine deaminase.

By "Adenosine Base Editor (ABE) polynucleotide" is meant a polynucleotide encoding an ABE. By "Adenosine Base Editor 8 (ABE8) polypeptide" or "ABE8" is meant a base editor as defined herein comprising an adenosine deaminase or adenosine deaminase variant comprising one or more of the alterations listed in **Table 14**, one of the combinations of alterations listed in **Table 14**, or an alteration at one or more of the amino acid positions listed in **Table 14**, such alterations are relative to the following reference sequence of the following reference sequence:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH
RVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 1), or a corresponding position in another adenosine deaminase. In embodiments, ABE8 comprises alterations at amino acids 82 and/or 166 of SEQ ID NO: 1.

In some embodiments, ABE8 comprises further alterations, as described herein, relative to the reference sequence.

By "Adenosine Base Editor 8 (ABE8) polynucleotide" is meant a polynucleotide encoding an ABE8 polypeptide.

"Administering" is referred to herein as providing one or more compositions described herein to a patient or a subject.

By "agent" is meant any small molecule chemical compound, antibody, nucleic acid molecule, or polypeptide, or fragments thereof.

By "alteration" is meant a change (increase or decrease) in the level, structure, or activity of an analyte, gene or polypeptide as detected by standard art known methods such as those described herein. As used herein, an alteration includes a 10% change in expression levels, a 25% change, a 40% change, and a 50% or greater change in expression levels. In some embodiments, an alteration includes an insertion, deletion, or substitution of a nucleobase or amino acid.

By "ameliorate" is meant decrease, suppress, attenuate, diminish, arrest, or stabilize the development or progression of a disease.

By "analog" is meant a molecule that is not identical, but has analogous functional or structural features. For example, a polypeptide analog retains the biological activity of a corresponding naturally-occurring polypeptide, while having certain biochemical modifications that enhance the analog's function relative to a naturally occurring polypeptide. Such biochemical modifications could increase the analog's protease resistance, membrane permeability, or half-life, without altering, for example, ligand binding. An analog may include an unnatural amino acid.

By "base editor (BE)," or "nucleobase editor polypeptide (NBE)" is meant an agent that binds a polynucleotide and has nucleobase modifying activity. In various embodiments, the base editor comprises a nucleobase modifying polypeptide (*e.g.*, a deaminase) and a polynucleotide programmable nucleotide binding domain (*e.g.*, Cas9 or Cpf1) in conjunction with a guide polynucleotide (*e.g.*, guide RNA (gRNA)). Representative nucleic acid and protein sequences of base editors are provided in the Sequence Listing as SEQ ID NOs: 2-11.

By "base editing activity" is meant acting to chemically alter a base within a polynucleotide. In one embodiment, a first base is converted to a second base. In one embodiment, the base editing activity is cytidine deaminase activity, *e.g.*, converting target C•G to T•A. In another embodiment, the base editing activity is adenosine or adenine deaminase activity, *e.g.*, converting A•T to G•C.

The term "base editor system" refers to an intermolecular complex for editing a nucleobase of a target nucleotide sequence. In various embodiments, the base editor (BE) system comprises (1) a polynucleotide programmable nucleotide binding domain, a deaminase domain (*e.g.*, cytidine deaminase or adenosine deaminase) for deaminating nucleobases in the

target nucleotide sequence; and (2) one or more guide polynucleotides (*e.g.*, guide RNA) in conjunction with the polynucleotide programmable nucleotide binding domain. In various embodiments, the base editor (BE) system comprises a nucleobase editor domain selected from an adenosine deaminase or a cytidine deaminase, and a domain having nucleic acid sequence specific binding activity. In some embodiments, the base editor system comprises (1) a base editor (BE) comprising a polynucleotide programmable DNA binding domain and a deaminase domain for deaminating one or more nucleobases in a target nucleotide sequence; and (2) one or more guide RNAs in conjunction with the polynucleotide programmable DNA binding domain. In some embodiments, the polynucleotide programmable nucleotide binding domain is a polynucleotide programmable DNA binding domain. In some embodiments, the base editor is a cytidine base editor (CBE). In some embodiments, the base editor is an adenine or adenosine base editor (ABE). In some embodiments, the base editor is an adenine or adenosine base editor (ABE) or a cytidine or a cytosine base editor (CBE).

The term "Cas9" or "Cas9 domain" refers to an RNA guided nuclease comprising a Cas9 protein, or a fragment thereof (*e.g.*, a protein comprising an active, inactive, or partially active DNA cleavage domain of Cas9, and/or the gRNA binding domain of Cas9). A Cas9 nuclease is also referred to sometimes as a casnl nuclease or a CRISPR (clustered regularly interspaced short palindromic repeat) associated nuclease.

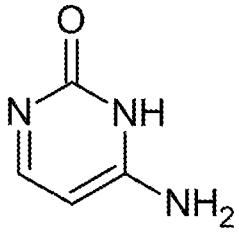
The term "conservative amino acid substitution" or "conservative mutation" refers to the replacement of one amino acid by another amino acid with a common property. A functional way to define common properties between individual amino acids is to analyze the normalized frequencies of amino acid changes between corresponding proteins of homologous organisms (Schulz, G. E. and Schirmer, R. H., Principles of Protein Structure, Springer-Verlag, New York (1979)). According to such analyses, groups of amino acids can be defined where amino acids within a group exchange preferentially with each other, and therefore resemble each other most in their impact on the overall protein structure (Schulz, G. E. and Schirmer, R. H., supra). Non-limiting examples of conservative mutations include amino acid substitutions of amino acids, for example, lysine for arginine and vice versa such that a positive charge can be maintained; glutamic acid for aspartic acid and vice versa such that a negative charge can be maintained; serine for threonine such that a free –OH can be maintained; and glutamine for asparagine such that a free $–NH_2$ can be maintained.

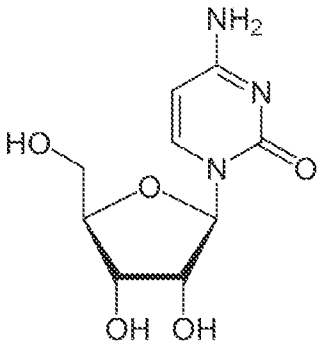
The term "coding sequence" or "protein coding sequence" as used interchangeably herein refers to a segment of a polynucleotide that codes for a protein. Coding sequences can also be referred to as open reading frames. The region or sequence is bounded nearer the 5' end by a

start codon and nearer the 3' end with a stop codon. Stop codons useful with the base editors described herein include the following:

Glutamine	**C**AG → **T**AG	Stop codon
	CAA → **T**AA	
Arginine	CGA → **T**GA	
Tryptophan	TG**G** → TG**A**	
	T**G**G → T**A**G	
	T**GG** → T**AA**	

By "complex" is meant a combination of two or more molecules whose interaction relies on inter-molecular forces. Non-limiting examples of inter-molecular forces include covalent and non-covalent interactions. Non-limiting examples of non-covalent interactions include hydrogen bonding, ionic bonding, halogen bonding, hydrophobic bonding, van der Waals interactions (e.g., dipole-dipole interactions, dipole-induced dipole interactions, and London dispersion forces), and π-effects. In an embodiment, a complex comprises polypeptides, polynucleotides, or a combination of one or more polypeptides and one or more polynucleotides. In one embodiment, a complex comprises one or more polypeptides that associate to form a base editor (e.g., base editor comprising a nucleic acid programmable DNA binding protein, such as Cas9, and a deaminase) and a polynucleotide (e.g., a guide RNA). In an embodiment, the complex is held together by hydrogen bonds. It should be appreciated that one or more components of a base editor (e.g., a deaminase, or a nucleic acid programmable DNA binding protein) may associate covalently or non covalently. As one example, a base editor may include a deaminase covalently linked to a nucleic acid programmable DNA binding protein (e.g., by a peptide bond). Alternatively, a base editor may include a deaminase and a nucleic acid programmable DNA binding protein that associate noncovalently (e.g., where one or more components of the base editor are supplied in trans and associate directly or via another molecule such as a protein or nucleic acid). In an embodiment, one or more components of the complex are held together by hydrogen bonds. Throughout the present disclosure, wherever an embodiment of a base editor is contemplated as containing a fusion protein, complexes comprising one or more domains of the base editor, or fragments thereof, are also contemplated.

By "cytosine" or " 4-Aminopyrimidin-2(1*H*)-one" is meant a purine nucleobase with the molecular formula $C_4H_5N_3O$, having the structure , and corresponding to CAS No. 71-30-7.

By "cytidine" is meant a cytosine molecule attached to a ribose sugar via a glycosidic bond, having the structure , and corresponding to CAS No. 65-46-3. Its molecular formula is $C_9H_{13}N_3O_5$.

By "Cytidine Base Editor (CBE)" is meant a base editor comprising a cytidine deaminase.

By "Cytidine Base Editor (CBE) polynucleotide" is meant a polynucleotide comprising a CBE.

By "cytidine deaminase" or "cytosine deaminase" is meant a polypeptide or fragment thereof capable of deaminating cytidine or cytosine. In one embodiment, the cytidine deaminase converts cytosine to uracil or 5-methylcytosine to thymine. The terms "cytidine deaminase" and "cytosine deaminase" are used interchangeably throughout the application. *Petromyzon marinus* cytosine deaminase 1 (PmCDA1) (SEQ ID NO: 12-13), Activation-induced cytidine deaminase (AICDA) (SEQ ID NOs: 14-16, and 18-21), and APOBEC (SEQ ID NOs: 22-62) are expemplary cytidine deaminases. Further exemplary cytidine deaminase (CDA) sequences are provided in the Sequence Listing as SEQ ID NOs: 63-67 and SEQ ID NOs: 68-190.

By "cytosine deaminase activity" is meant catalyzing the deamination of cytosine or cytidine. In one embodiment, a polypeptide having cytosine deaminase activity converts an amino group to a carbonyl group. In an embodiment, a cytosine deaminase converts cytosine to uracil (i.e., C to U) or 5-methylcytosine to thymine (i.e., 5mC to T). In some embodiments, a cytosine deaminase variant as provided herein has increased cytosine deaminase activity (*e.g.*, at least 10-fold, 20-fold, 30-fold, 40-fold, 50-fold, 60-fold, 70-fold, 80-fold, 90-fold, 100-fold or

more) relative to a reference cytosine deaminase. The term "deaminase" or "deaminase domain," as used herein, refers to a protein or fragment thereof that catalyzes a deamination reaction.

"Detect" refers to identifying the presence, absence or amount of the analyte to be detected. In one embodiment, a sequence alteration in a polynucleotide or polypeptide is detected. In another embodiment, the presence of indels is detected.

By "detectable label" is meant a composition that when linked to a molecule of interest renders the latter detectable, via spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive isotopes, magnetic beads, metallic beads, colloidal particles, fluorescent dyes, electron-dense reagents, enzymes (for example, as commonly used in an enzyme linked immunosorbent assay (ELISA)), biotin, digoxigenin, or haptens.

By "disease" is meant any condition or disorder that damages or interferes with the normal function of a cell, tissue, or organ.

By "effective amount" is meant the amount of an agent or active compound, *e.g.*, a base editor as described herein, that is required to ameliorate the symptoms of a disease relative to an untreated patient or an individual without disease, *i.e.*, a healthy individual, or is the amount of the agent or active compound sufficient to elicit a desired biological response. The effective amount of active compound(s) used to practice the present invention for therapeutic treatment of a disease varies depending upon the manner of administration, the age, body weight, and general health of the subject. Ultimately, the attending physician or veterinarian will decide the appropriate amount and dosage regimen. Such amount is referred to as an "effective" amount. In one embodiment, an effective amount is the amount of a base editor of the invention sufficient to introduce an alteration in a gene of interest in a cell (*e.g.*, a cell *in vitro* or *in vivo*). In one embodiment, an effective amount is the amount of a base editor required to achieve a therapeutic effect. Such therapeutic effect need not be sufficient to alter a pathogenic gene in all cells of a subject, tissue or organ, but only to alter the pathogenic gene in about 1%, 5%, 10%, 25%, 50%, 75% or more of the cells present in a subject, tissue or organ. In one embodiment, an effective amount is sufficient to ameliorate one or more symptoms of a disease.

The term "exonuclease" refers to a protein or polypeptide capable of digesting a nucleic acid (*e.g.*, RNA or DNA) from free ends.

The term "endonuclease" refers to a protein or polypeptide capable of catalyzing (*e.g.*, cleaving) internal regions in a nucleic acid (*e.g.*, DNA or RNA).

By "fragment" is meant a portion of a polypeptide or nucleic acid molecule. This portion contains, at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% of the entire length of the reference nucleic acid molecule or polypeptide. A fragment may contain 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 nucleotides or amino acids.

By "guide polynucleotide" is meant a polynucleotide or polynucleotide complex which is specific for a target sequence and can form a complex with a polynucleotide programmable nucleotide binding domain protein (*e.g.*, Cas9 or Cpf1). In an embodiment, the guide polynucleotide is a guide RNA (gRNA). gRNAs can exist as a complex of two or more RNAs, or as a single RNA molecule.

In some embodiments the guide polynucleotide has a nucleotide sequence selected from the following, where a lowercase "g" indicates a 5' mismatch to the target sequence:

a) gGUUUUAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 191);

b) gUUUCUUACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 192);

c) gGUUUCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 193);

d) GCCACUUACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 194);

e) gACAUUAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 195);

f) gGAUCUCACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 196);

g) gUCCUUAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 197);

h) GUCACCUACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 198);

i) GAUUUCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 190);

j) gGUGCUUACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 200);

k) gUCCACAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 201);

l) GAUACUUACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 202);

m) gUGUUUUAGCUGCGGCAAGGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 203);

n) gUUUCUUACAGCCAUAAUUUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 204);

o) gCUCCACAGCUGCGGCAAGGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 205);

p) GAUACUUACAGCCAUAAUUUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 206);

q) gUGUUUUAGGGACGAAAGAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 207);

r) gUUACCUGGCUCUCUUAGCCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 208);

s) gCUCCACAGGGACGAAAGAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUC CGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 209);

t) gCUUGCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 210);

u) gAUUGCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 211);

v) gUCUCCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 212);

w) gUCUGCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 213);

x) gGACUCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 214);

y) GCACCCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 215);

z) gAAUUUAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 216);

aa) gCAUUAGGUCGAGAUCACAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 217);

bb) gCCUUAGGUCGAGAUCACAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 218);

cc) GUUUCAGGUCGAGAUCACAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 219);

dd) gACAUUAGGCUAAGAGAGCCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 220);

ee) gUCCUUAGGCUAAGAGAGCCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 221);

ff) gGUUUCAGGCUAAGAGAGCCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 222);

gg) gACAUUAGAUUAUGGCUCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 223);

hh) gUCCUUAGAUUAUGGCUCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 224); or

ii) gGUUUCAGAUUAUGGCUCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 225).

By "heterologous," or "exogenous" is meant a polynucleotide or polypeptide that 1) has been experimentally incorporated to a polynucleotide or polypeptide sequence to which the polynucleotide or polypeptide is not normally found in nature; or 2) has been experimentally placed into a cell that does not normally comprise the polynucleotide or polypeptide. In some embodiments, "heterologous" means that a polynucleotide or polypeptide has been experimentally placed into a non-native context. In some embodiments, a heterologous polynucleotide or polypeptide is derived from a first species or host organism, and is incorporated into a polynucleotide or polypeptide derived from a second species or host organism. In some embodiments, the first species or host organism is different from the second species or host organism. In some embodiments the heterologous polynucleotide is DNA. In some embodiments the heterologous polynucleotide is RNA.

In some embodiments, a heterologous polynucleotide is a heterologous intron. In some embodiments, a heterologous intron is a synthetic intron. In some embodiments, a heterologous intron is derived from a mammalian gene (*e.g.*, NF1, PAX2, EEF1A1, HBB, IGHG1, SLC50A1, ABCB11, BRSK2, PLXNB3, TMPRSS6, IL32, ANTXRL, PKHD1L1, PADI1, KRT6C, or

HMCN2). In some embodiments, a heterologous intron is derived from a non-mammalian gene (*e.g.*, HMCN2-Salmon, ENPEP-Gecko). In some embodiments, polynucleotides encoding a base editor as provided herein include a heterologous intron. In some embodiments, the base editor is an adenosine base editor (ABE). In some embodiments, the base editor is a cytidine base editor (CBE).

In some embodiments, a heterologous intron is incorporated into a polynucleotide encoding a polynucleotide programmable DNA binding protein or fragment thereof. In some embodiments, the polynucleotide programmable DNA binding protein is a Cas9, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ domain. In some embodiments, the polynucleotide programmable DNA binding domain is a *Staphylococcus aureus* Cas9 (SaCas9), *Streptococcus thermophilus 1* Cas9 (St1Cas9), a *Streptococcus pyogenes* Cas9 (SpCas9), or variants thereof.

In some embodiments, a heterologous intron is incorporated into a polynucleotide encoding a deaminase or fragment thereof. In some embodiments, a heterologous intron is incorporated into a polynucleotide encoding an adenosine deaminase. In some embodiments, the adenosine deaminase is TadA. In some embodiments, a heterologous intron is incorporated into a polynucleotide encoding an cytidine deaminase.

"Hybridization" means hydrogen bonding, which may be Watson-Crick, Hoogsteen or reversed Hoogsteen hydrogen bonding, between complementary nucleobases. For example, adenine and thymine are complementary nucleobases that pair through the formation of hydrogen bonds.

By "increases" is meant a positive alteration of at least 10%, 25%, 50%, 75%, or 100%.

The terms "inhibitor of base repair", "base repair inhibitor", "IBR" or their grammatical equivalents refer to a protein that is capable in inhibiting the activity of a nucleic acid repair enzyme, for example a base excision repair enzyme.

An "intein" is a fragment of a protein that is able to excise itself and join the remaining fragments (the exteins) with a peptide bond in a process known as protein splicing.

By "intron" is meant a non-coding nucleotide sequence that is removed by splicing before translation of a transcript. In some embodiments, an intron is removed during the precursor messenger RNA stage of maturation of mRNA by RNA splicing. In some embodiments, an intron is derived from a gene of an organism. In some embodiments, an intron is synthetic. In some embodiments, an intron includes a splice acceptor and a splice donor site. In some embodiments, an intron is about 10, 25, 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450,

or 500 nucleotides in length. In some embodiments, an intron is about 50, 100, 125, 150, 175, or 200 nucleotides in length. In some embodiments, an intron is about 150 nucleotides in length.

In some embodiments, an intron is derived from a mammalian gene (*e.g.*, NF1, PAX2, EEF1A1, HBB, IGHG1, SLC50A1, ABCB11, BRSK2, PLXNB3, TMPRSS6, IL32, ANTXRL, PKHD1L1, PADI1, KRT6C, or HMCN2). In some embodiments, an intron is derived from a non-mammalian gene (*e.g.*, HMCN2-Salmon, ENPEP-Gecko). In some embodiments, an intron has a polynucleotide sequence selected from the following:

a) GTGAGATCAAATGAAAGTTTCATATAGAAATACAAAACCTAGAGAACTGGCATGTAAGAGAA GCAAAAATTACTTCAGCAAGGCCATGTTAGTAAATTTGCATCTGTTTGTCCACATTAG (SEQ ID NO: 226);

b) GTAGGTGACAATGCTGCAGCTGCCTAATCTAGGTGGGGGGAACTAAATTGTGGGTGAGCTGC TGAATGGTCTGTAGTCTGAGGCTGGGGTGGGGGGAGACACAACGTCCCCTCCCTGCAAACCA CTGCTATTCTGTCCCTCTCTCTCCTTAG (SEQ ID NO: 227);

c) GTAAGTGGCTTTCAAGACCATTGTTAAAAAGCTCTGGGAATGGCGATTTCATGCTTACATAA ATTGGCATGCTTGTGTTTCAG (SEQ ID NO: 228);

d) GTAAGTATCAAGGTTACAAGACAGGTTTAAGGAGACCAATAGAAACTGGGCTTGTCTAGACA GAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTTC TCTCCACAG (SEQ ID NO: 229);

e) GTAAGCACAACTGGGATGGGGTGACAGGGGTGCAAGATTGAAAACTGGCTCCTCTCCTCATA GCAGTTCTTGTGATTTCAG (SEQ ID NO: 230);

f) GTAAGAAATGTTATTTTTCAGTAAGTGATTTAGTTATTTTTCCTTTTTTCTCATTAAAATTT CTCTAACATCTCCCTCTTCATGTTTTAG (SEQ ID NO: 231);

g) GTGAGACCCTAGCCCCCTCAACCCTGCCCTGGCCTCTCCCCAAACCTGCCCCCCCACGCTGA CCCCCACACCCGGCCGCCCGCAG (SEQ ID NO: 232);

h) GTGGGTGTCAGAGGCATCGGGGCTGCGGGGTAGGGGGCTGCCCCACCCCTAACGAAGTCTGC TCCTCCAG (SEQ ID NO: 233);

i) GCAGGGAAGTCCTGCTTCCGTGCCCCACCGGTGCTCAGCTGAGGCTCCCTTGAAAATGCGAG GCTGTTTCCAACTTTGGTCTGTTTCCCTGGCAG (SEQ ID NO: 234);

j) GTGGGGAGTTGGGGTCCCCGAAGGTGAGGACCCTCTGGGGATGAGGGTGCTTCTCTGAGACA CTTTCTTTTCCTCACACCTGTTCCTCGCCAGCAG (SEQ ID NO: 235);

k) GTATAGACCCCTTGATCTCCTAACCCTAACCCTAACCCTAACCCTAACCTACAAAATCTTAG AGCATCAGTGGGAGCATCTCACTGTCCAGGCTCAATATTTCTTCATTTTCTTGCAG (SEQ ID NO: 236);

l) GTAATTATGATTAAAGATGGTGATTGTTTATTTTCTTTTATGATTGTCCTTAGTATTATGTA ACCTGCAAATTCTATTGCAG (SEQ ID NO: 237);

m) GTGAGTGACACAAGGTGTTGTCTGGGGAGTGGGGAAGGGGGATGGAAGTGAATCCTGTTGGT GGGGTGGAGAAAGGGCGATCTCAAGAGGGCCACTCTCTCCAG (SEQ ID NO: 238);

n) GTAAGCATCTCCACCATCCTTCTGTTTACTCTGATGGGGTCTGCAAAGGGGAGATGATGTAT AGGGTTGGGTATCTCTGTAAATGTCAGATGTGAAGTTGATCTTATGACCTTCTGTTCTGCAG (SEQ ID NO: 239);

o) GTGAGGGTCTCCCAGGCTGGGCAGGGGGAGGGGGCTGCTGCCTTGATTGCGTCCCAGGACAC AGCCCTCCTCCAGCCTGCCCTCGCCTTGCTCATCCCCTCCCCATCTCAGCCCCACCCCCACT AACTCTCTCTCTGCTCTGACTCAG (SEQ ID NO: 240);

p) GTAATGATTGATTGCAATGTATGATTACAATAATCTCAGTATAAGTTCAGTAATAATAACCT TCCACTGCTGTCCTCTGTGTGCACCCAG (SEQ ID NO: 241); or

q) GTAAATATATACAACAGTTTTTCATTTAAATAAGTGCACGGCACAAATAAGAAAAATATGTC AAAAATGTAACCAATAGTTTTTTTCAAATTTAG (SEQ ID NO: 242).

In some embodiments, polynucleotides encoding a base editor as provided herein include a heterologous intron. In some embodiments, the base editor is an adenosine base editor (ABE). In some embodiments, the base editor is a cytidine base editor (CBE). In some embodiments, an intron is heterologously incorporated into a polynucleotide sequence. In some embodiments, the polynucleotide sequence is DNA. In some embodiments, the polynucleotide sequence is RNA. In some embodiments, an intron is heterologously incorporated into a polynucleotide encoding a polynucleotide programmable DNA binding protein. In some embodiments, the polynucleotide programmable DNA binding protein is a Cas9, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ domain. In some embodiments, the polynucleotide programmable DNA binding domain is a *Staphylococcus aureus* Cas9 (SaCas9), *Streptococcus thermophilus 1* Cas9 (St1Cas9), a *Streptococcus pyogenes* Cas9 (SpCas9), or variants thereof.

In some embodiments, an intron is heterologously incorporated into a polynucleotide encoding a deaminase. In some embodiments, an intron is heterologously incorporated into a polynucleotide encoding an adenosine deaminase. In some embodiments, the adenosine deaminase is TadA. In some embodiments, an intron is heterologously incorporated into a

polynucleotide encoding a cytidine deaminase. In some embodiments, an intron is heterologously incorporated into a polynucleotide programmable DNA binding protein (*e.g.*, Cas9). In some embodiments, an intron is heterologously incorporated into a linker region.

The terms "isolated," "purified," or "biologically pure" refer to material that is free to varying degrees from components which normally accompany it as found in its native state. "Isolate" denotes a degree of separation from original source or surroundings. "Purify" denotes a degree of separation that is higher than isolation. A "purified" or "biologically pure" protein is sufficiently free of other materials such that any impurities do not materially affect the biological properties of the protein or cause other adverse consequences. That is, a nucleic acid or peptide of this invention is purified if it is substantially free of cellular material, viral material, or culture medium when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. Purity and homogeneity are typically determined using analytical chemistry techniques, for example, polyacrylamide gel electrophoresis or high performance liquid chromatography. The term "purified" can denote that a nucleic acid or protein gives rise to essentially one band in an electrophoretic gel. For a protein that can be subjected to modifications, for example, phosphorylation or glycosylation, different modifications may give rise to different isolated proteins, which can be separately purified.

By "isolated polynucleotide" is meant a nucleic acid molecule that is free of the genes which, in the naturally-occurring genome of the organism from which the nucleic acid molecule of the invention is derived, flank the gene. In embodiments, the nucleic acid molecule contains DNA or is a DNA molecule. The term therefore includes, for example, a recombinant DNA that is incorporated into a vector; into an autonomously replicating plasmid or virus; or into the genomic DNA of a prokaryote or eukaryote; or that exists as a separate molecule (for example, a cDNA or a genomic or cDNA fragment produced by PCR or restriction endonuclease digestion) independent of other sequences. In addition, the term includes an RNA molecule that is transcribed from a DNA molecule, as well as a recombinant DNA that is part of a hybrid gene encoding additional polypeptide sequence.

By an "isolated polypeptide" is meant a polypeptide of the invention that has been separated from components that naturally accompany it. Typically, the polypeptide is isolated when it is at least 60%, by weight, free from the proteins and naturally-occurring organic molecules with which it is naturally associated. Preferably, the preparation is at least 75%, more preferably at least 90%, and most preferably at least 99%, by weight, a polypeptide of the invention. An isolated polypeptide of the invention may be obtained, for example, by extraction from a natural source, by expression of a recombinant nucleic acid encoding such a polypeptide;

or by chemically synthesizing the protein. Purity can be measured by any appropriate method, for example, column chromatography, polyacrylamide gel electrophoresis, or by HPLC analysis.

The term "linker", as used herein, refers to a molecule that links two moieties. In one embodiment, the term "linker" refers to a covalent linker (*e.g.*, covalent bond) or a non-covalent linker.

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

The terms "nucleic acid" and "nucleic acid molecule," as used herein, refer to a compound comprising a nucleobase and an acidic moiety, *e.g.*, a nucleoside, a nucleotide, or a polymer of nucleotides. Typically, polymeric nucleic acids, *e.g.*, nucleic acid molecules comprising three or more nucleotides are linear molecules, in which adjacent nucleotides are linked to each other via a phosphodiester linkage. In some embodiments, "nucleic acid" refers to individual nucleic acid residues (*e.g.* nucleotides and/or nucleosides). In some embodiments, "nucleic acid" refers to an oligonucleotide chain comprising three or more individual nucleotide residues. As used herein, the terms "oligonucleotide" and "polynucleotide" can be used interchangeably to refer to a polymer of nucleotides (*e.g.*, a string of at least three nucleotides). In some embodiments, "nucleic acid" encompasses RNA as well as single and/or double-stranded DNA. Nucleic acids may be naturally occurring, for example, in the context of a genome, a transcript, an mRNA, tRNA, rRNA, siRNA, snRNA, a plasmid, cosmid, chromosome, chromatid, or other naturally occurring nucleic acid molecule. On the other hand, a nucleic acid molecule may be a non-naturally occurring molecule, *e.g.*, a recombinant DNA or RNA, an artificial chromosome, an engineered genome, or fragment thereof, or a synthetic DNA, RNA, DNA/RNA hybrid, or including non-naturally occurring nucleotides or nucleosides.

Furthermore, the terms "nucleic acid," "DNA," "RNA," and/or similar terms include nucleic acid analogs, *e.g.,* analogs having other than a phosphodiester backbone. Nucleic acids can be purified from natural sources, produced using recombinant expression systems and optionally purified, chemically synthesized, etc. Where appropriate, *e.g.*, in the case of chemically synthesized molecules, nucleic acids can comprise nucleoside analogs such as

analogs having chemically modified bases or sugars, and backbone modifications. A nucleic acid sequence is presented in the 5′ to 3′ direction unless otherwise indicated. In some embodiments, a nucleic acid is or comprises natural nucleosides (*e.g.* adenosine, thymidine, guanosine, cytidine, uridine, deoxyadenosine, deoxythymidine, deoxyguanosine, and deoxycytidine); nucleoside analogs (*e.g.*, 2-aminoadenosine, 2-thiothymidine, inosine, pyrrolo-pyrimidine, 3-methyl adenosine, 5-methylcytidine, 2-aminoadenosine, C5-bromouridine, C5-fluorouridine, C5-iodouridine, C5-propynyl-uridine, C5-propynyl-cytidine, C5-methylcytidine, 2-aminoadenosine, 7-deazaadenosine, 7-deazaguanosine, 8-oxoadenosine, 8-oxoguanosine, O(6)-methylguanine, and 2-thiocytidine); chemically modified bases; biologically modified bases (*e.g.*, methylated bases); intercalated bases; modified sugars (2′-*e.g.*, fluororibose, ribose, 2′-deoxyribose, arabinose, and hexose); and/or modified phosphate groups (*e.g.*, phosphorothioates and 5′-*N*-phosphoramidite linkages).

The term "nuclear localization sequence," "nuclear localization signal," or "NLS" refers to an amino acid sequence that promotes import of a protein into the cell nucleus. Nuclear localization sequences are known in the art and described, for example, in Plank *et al.*, International PCT application, PCT/EP2000/011690, filed November 23, 2000, published as WO/2001/038547 on May 31, 2001, the contents of which are incorporated herein by reference for their disclosure of exemplary nuclear localization sequences. In other embodiments, the NLS is an optimized NLS described, for example, by Koblan *et al.*, Nature Biotech. 2018 doi:10.1038/nbt.4172. In some embodiments, an NLS comprises the amino acid sequence KRTADGSEFESPKKKRKV (SEQ ID NO: 243), KRPAATKKAGQAKKKK (SEQ ID NO: 244), KKTELQTTNAENKTKKL (SEQ ID NO: 245), KRGINDRNFWRGENGRKTR (SEQ ID NO: 246), RKSGKIAAIVVKRPRK (SEQ ID NO: 247), PKKKRKV (SEQ ID NO: 248), or MDSLLMNRRKFLYQFKNVRWAKGRRETYLC (SEQ ID NO: 249).

The term "nucleobase," "nitrogenous base," or "base," used interchangeably herein, refers to a nitrogen-containing biological compound that forms a nucleoside, which in turn is a component of a nucleotide. The ability of nucleobases to form base pairs and to stack one upon another leads directly to long-chain helical structures such as ribonucleic acid (RNA) and deoxyribonucleic acid (DNA). Five nucleobases – adenine (A), cytosine (C), guanine (G), thymine (T), and uracil (U) – are called primary or canonical. Adenine and guanine are derived from purine, and cytosine, uracil, and thymine are derived from pyrimidine. DNA and RNA can also contain other (non-primary) bases that are modified. Non-limiting exemplary modified nucleobases can include hypoxanthine, xanthine, 7-methylguanine, 5,6-dihydrouracil, 5-methylcytosine (m5C), and 5-hydromethylcytosine. Hypoxanthine and xanthine can be created

through mutagen presence, both of them through deamination (replacement of the amine group with a carbonyl group). Hypoxanthine can be modified from adenine. Xanthine can be modified from guanine. Uracil can result from deamination of cytosine. A "nucleoside" consists of a nucleobase and a five carbon sugar (either ribose or deoxyribose). Examples of a nucleoside include adenosine, guanosine, uridine, cytidine, 5-methyluridine (m5U), deoxyadenosine, deoxyguanosine, thymidine, deoxyuridine, and deoxycytidine. Examples of a nucleoside with a modified nucleobase includes inosine (I), xanthosine (X), 7-methylguanosine (m7G), dihydrouridine (D), 5-methylcytidine (m5C), and pseudouridine (Ψ). A "nucleotide" consists of a nucleobase, a five carbon sugar (either ribose or deoxyribose), and at least one phosphate group. Non-limiting examples of modified nucleobases and/or chemical modifications that a modified nucleobase may include are the following: pseudo-uridine, 5-Methyl-cytosine, 2′-O-methyl-3′-phosphonoacetate, 2′-*O*-methyl thioPACE (MSP), 2′-*O*-methyl-PACE (MP), 2′-fluoro RNA (2′-F-RNA), constrained ethyl (S-cEt), 2′-O-methyl ('M'), 2′-O-methyl-3′-phosphorothioate ('MS'), 2′-O-methyl-3′-thiophosphonoacetate ('MSP'), 5-methoxyuridine, phosphorothioate, and N1-Methylpseudouridine.

The term "nucleic acid programmable DNA binding protein" or "napDNAbp" may be used interchangeably with "polynucleotide programmable nucleotide binding domain" to refer to a protein that associates with a nucleic acid (*e.g.*, DNA or RNA), such as a guide nucleic acid or guide polynucleotide (*e.g.,* gRNA), that guides the napDNAbp to a specific nucleic acid sequence. In some embodiments, the polynucleotide programmable nucleotide binding domain is a polynucleotide programmable DNA binding domain. In some embodiments, the polynucleotide programmable nucleotide binding domain is a polynucleotide programmable RNA binding domain. In some embodiments, the polynucleotide programmable nucleotide binding domain is a Cas9 protein. A Cas9 protein can associate with a guide RNA that guides the Cas9 protein to a specific DNA sequence that is complementary to the guide RNA. In some embodiments, the napDNAbp is a Cas9 domain, for example a nuclease active Cas9, a Cas9 nickase (nCas9), or a nuclease inactive Cas9 (dCas9). Non-limiting examples of nucleic acid programmable DNA binding proteins include, Cas9 (*e.g.*, dCas9 and nCas9), Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, and Cas12j/CasΦ (Cas12j/Casphi). Non-limiting examples of Cas enzymes include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5d, Cas5t, Cas5h, Cas5a, Cas6, Cas7, Cas8, Cas8a, Cas8b, Cas8c, Cas9 (also known as Csn1 or Csx12), Cas10, Cas10d, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, Cas12j/CasΦ, Cpf1, Csy1 , Csy2, Csy3, Csy4, Cse1, Cse2, Cse3, Cse4, Cse5e, Csc1, Csc2, Csa5, Csn1, Csn2, Csm1, Csm2, Csm3,

Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx1S, Csx11, Csf1, Csf2, CsO, Csf4, Csd1, Csd2, Cst1, Cst2, Csh1, Csh2, Csa1, Csa2, Csa3, Csa4, Csa5, Type II Cas effector proteins, Type V Cas effector proteins, Type VI Cas effector proteins, CARF, DinG, homologues thereof, or modified or engineered versions thereof. Other nucleic acid programmable DNA binding proteins are also within the scope of this disclosure, although they may not be specifically listed in this disclosure. See, *e.g.*, Makarova *et al.* "Classification and Nomenclature of CRISPR-Cas Systems: Where from Here?" *CRISPR J.* 2018 Oct;1:325-336. doi: 10.1089/crispr.2018.0033; Yan *et al.*, "Functionally diverse type V CRISPR-Cas systems" *Science*. 2019 Jan 4;363(6422):88-91. doi: 10.1126/science.aav7271, the entire contents of each are hereby incorporated by reference. Exemplary nucleic acid programmable DNA binding proteins and nucleic acid sequences encoding nucleic acid programmable DNA binding proteins are provided in the Sequence Listing as SEQ ID NOs: 250-283 and 490.

The terms "nucleobase editing domain" or "nucleobase editing protein," as used herein, refers to a protein or enzyme that can catalyze a nucleobase modification in RNA or DNA, such as cytosine (or cytidine) to uracil (or uridine) or thymine (or thymidine), and adenine (or adenosine) to hypoxanthine (or inosine) deaminations, as well as non-templated nucleotide additions and insertions. In some embodiments, the nucleobase editing domain is a deaminase domain (*e.g.*, an adenine deaminase or an adenosine deaminase; or a cytidine deaminase or a cytosine deaminase).

As used herein, "obtaining" as in "obtaining an agent" includes synthesizing, purchasing, or otherwise acquiring the agent.

A "patient" or "subject" as used herein refers to a mammalian subject or individual diagnosed with, at risk of having or developing, or suspected of having or developing a disease or a disorder. In some embodiments, the term "patient" refers to a mammalian subject with a higher than average likelihood of developing a disease or a disorder. Exemplary patients can be humans, non-human primates, cats, dogs, pigs, cattle, cats, horses, camels, llamas, goats, sheep, rodents (*e.g.*, mice, rabbits, rats, or guinea pigs) and other mammalians that can benefit from the therapies disclosed herein. Exemplary human patients can be male and/or female.

"Patient in need thereof" or "subject in need thereof" is referred to herein as a patient diagnosed with, at risk or having, predetermined to have, or suspected of having a disease or disorder.

The terms "pathogenic mutation", "pathogenic variant", "disease causing mutation", "disease causing variant", "deleterious mutation", or "predisposing mutation" refers to a genetic

alteration or mutation that is associated with a disease or disorder or that increases an individual's susceptibility or predisposition to a certain disease or disorder. In some embodiments, the pathogenic mutation comprises at least one wild-type amino acid substituted by at least one pathogenic amino acid in a protein encoded by a gene. In some embodiments, the pathogenic mutation is in a terminating region (*e.g.*, stop codon). In some embodiments, the pathogenic mutation is in a non-coding region (*e.g.*, intron, promoter, etc.).

The terms "protein", "peptide", "polypeptide", and their grammatical equivalents are used interchangeably herein, and refer to a polymer of amino acid residues linked together by peptide (amide) bonds. A protein, peptide, or polypeptide can be naturally occurring, recombinant, or synthetic, or any combination thereof.

The term "fusion protein" as used herein refers to a hybrid polypeptide which comprises protein domains from at least two different proteins.

The term "recombinant" as used herein in the context of proteins or nucleic acids refers to proteins or nucleic acids that do not occur in nature, but are the product of human engineering. For example, in some embodiments, a recombinant protein or nucleic acid molecule comprises an amino acid or nucleotide sequence that comprises at least one, at least two, at least three, at least four, at least five, at least six, or at least seven mutations as compared to any naturally occurring sequence.

By "reduces" is meant a negative alteration of at least 10%, 25%, 50%, 75%, or 100%.

By "reference" is meant a standard or control condition. In one embodiment, the reference is the level of editing provided by a base editor encoded by a polynucleotide that does not include an intron. In another embodiment, the reference is the level of editing provided by a base editor encoded by a polynucleotide comprising an intron that does not include an alteration in a splice acceptor or splice donor site. In one embodiment, the reference is the level, structure or activity of an analyte present in a wild type or healthy cell. In other embodiments and without limitation, a reference is the level, structure or activity of an analyte present in an untreated cell that is not subjected to a test condition, or is subjected to placebo or normal saline, medium, buffer, and/or a control vector that does not harbor a polynucleotide of interest.

A "reference sequence" is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset of or the entirety of a specified sequence; for example, a segment of a full-length cDNA or gene sequence, or the complete cDNA or gene sequence. For polypeptides, the length of the reference polypeptide sequence will generally be at least about 16 amino acids, at least about 20 amino acids, at least about 25 amino acids, about 35 amino acids, about 50 amino acids, or about 100 amino acids. For nucleic acids, the length of the reference

nucleic acid sequence will generally be at least about 50 nucleotides, at least about 60 nucleotides, at least about 75 nucleotides, about 100 nucleotides or about 300 nucleotides or any integer thereabout or therebetween. In some embodiments, a reference sequence is a wild-type sequence of a protein of interest. In other embodiments, a reference sequence is a polynucleotide sequence encoding a wild-type protein.

The terms "RNA-programmable nuclease," and "RNA-guided nuclease" are used with (*e.g.*, binds or associates with) one or more RNA(s) that is not a target for cleavage. In some embodiments, an RNA-programmable nuclease, when in a complex with an RNA, may be referred to as a nuclease:RNA complex. Typically, the bound RNA(s) is referred to as a guide RNA (gRNA). In some embodiments, the RNA-programmable nuclease is the (CRISPR-associated system) Cas9 endonuclease, for example, Cas9 (Csnl) from *Streptococcus pyogenes*, (e.g., SEQ ID NO: 250), Cas9 from *Neisseria meningitidis* (NmeCas9; SEQ ID NO: 261), Nme2Cas9 (SEQ ID NO: 262), or derivatives thereof (e.g. a sequence with at least about 85% sequence identity to a Cas9, such as Nme2Cas9 or spCas9).

The term “single nucleotide polymorphism (SNP)” is a variation in a single nucleotide that occurs at a specific position in the genome, where each variation is present to some appreciable degree within a population (*e.g.*, > 1%).

By "specifically binds" is meant a nucleic acid molecule, polypeptide, polypeptide/polynucleotide complex, compound, or molecule that recognizes and binds a polypeptide and/or nucleic acid molecule of the invention, but which does not substantially recognize and bind other molecules in a sample, for example, a biological sample.

By "substantially identical" is meant a polypeptide or nucleic acid molecule exhibiting at least 50% identity to a reference amino acid sequence. In one embodiment, a reference sequence is a wild-type amino acid or nucleic acid sequence. In another embodiment, a reference sequence is any one of the amino acid or nucleic acid sequences described herein. In one embodiment, such a sequence is at least 60%, 80%, 85%, 90%, 95% or even 99% identical at the amino acid level or nucleic acid level to the sequence used for comparison.

Sequence identity is typically measured using sequence analysis software (for example, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705, BLAST, BESTFIT, GAP, or PILEUP/PRETTYBOX programs). Such software matches identical or similar sequences by assigning degrees of homology to various substitutions, deletions, and/or other modifications. Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid, asparagine,

glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine. In an exemplary approach to determining the degree of identity, a BLAST program may be used, with a probability score between e^{-3} and e^{-100} indicating a closely related sequence.

COBALT is used, for example, with the following parameters:

a) alignment parameters: Gap penalties-11,-1 and End-Gap penalties-5,-1,

b) CDD Parameters: Use RPS BLAST on; Blast E-value 0.003; Find Conserved columns and Recompute on, and

c) Query Clustering Parameters: Use query clusters on; Word Size 4; Max cluster distance 0.8; Alphabet Regular.

EMBOSS Needle is used, for example, with the following parameters:

a) Matrix: BLOSUM62;

b) GAP OPEN: 10;

c) GAP EXTEND: 0.5;

d) OUTPUT FORMAT: pair;

e) END GAP PENALTY: false;

f) END GAP OPEN: 10; and

g) END GAP EXTEND: 0.5.

Nucleic acid molecules useful in the methods of the invention include any nucleic acid molecule that encodes a polypeptide of the invention or a fragment thereof. Such nucleic acid molecules need not be 100% identical with an endogenous nucleic acid sequence, but will typically exhibit substantial identity. Polynucleotides having "substantial identity" to an endogenous sequence are typically capable of hybridizing with at least one strand of a double-stranded nucleic acid molecule. Nucleic acid molecules useful in the methods of the invention include any nucleic acid molecule that encodes a polypeptide of the invention or a fragment thereof. Such nucleic acid molecules need not be 100% identical with an endogenous nucleic acid sequence, but will typically exhibit substantial identity. Polynucleotides having "substantial identity" to an endogenous sequence are typically capable of hybridizing with at least one strand of a double-stranded nucleic acid molecule. By "hybridize" is meant pair to form a double-stranded molecule between complementary polynucleotide sequences (*e.g.*, a gene described herein), or portions thereof, under various conditions of stringency. (See, *e.g.*, Wahl, G. M. and S. L. Berger (1987) Methods Enzymol. 152:399; Kimmel, A. R. (1987) Methods Enzymol. 152:507).

For example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium

citrate, and more preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, *e.g.*, formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and more preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30° C, more preferably of at least about 37° C, and most preferably of at least about 42° C. Varying additional parameters, such as hybridization time, the concentration of detergent, *e.g.*, sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred: embodiment, hybridization will occur at 30° C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37° C in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 µg/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42° C in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50% formamide, and 200 µg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

For most applications, washing steps that follow hybridization will also vary in stringency. Wash stringency conditions can be defined by salt concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include a temperature of at least about 25° C, more preferably of at least about 42° C, and even more preferably of at least about 68° C. In an embodiment, wash steps will occur at 25° C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In another embodiment, wash steps will occur at 42 C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 68° C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Additional variations on these conditions will be readily apparent to those skilled in the art. Hybridization techniques are well known to those skilled in the art and are described, for example, in Benton and Davis (Science 196:180, 1977); Grunstein and Hogness (Proc. Natl. Acad. Sci., USA 72:3961, 1975); Ausubel *et al.* (Current Protocols in Molecular Biology, Wiley Interscience, New York, 2001); Berger and Kimmel (Guide to Molecular Cloning Techniques, 1987, Academic Press, New York); and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York.

By “split” is meant divided into two or more fragments.

A "split Cas9 protein" or "split Cas9" refers to a Cas9 protein that is provided as an N-terminal fragment and a C-terminal fragment encoded by two separate nucleotide sequences. The polypeptides corresponding to the N-terminal portion and the C-terminal portion of the Cas9 protein may be spliced to form a “reconstituted” Cas9 protein.

The term "target site" refers to a sequence within a nucleic acid molecule that is modified. In embodiments, the nucleic acid molecule is deaminated by a deaminase, a fusion protein or complex comprising a deaminase, or a base editor as disclosed herein. In embodiments, the deaminase is a cytidine or adenine deaminase. In some instances, the deaminase is a dCas9-adenosine deaminase fusion protein. In some cases, the base editor is an adenine or adenosine base editor (ABE) or a cytidine or a cytosine base editor (CBE).

As used herein, the terms “treat,” treating,” “treatment,” and the like refer to reducing or ameliorating a disorder and/or symptoms associated therewith or obtaining a desired pharmacologic and/or physiologic effect. It will be appreciated that, although not precluded, treating a disorder or condition does not require that the disorder, condition or symptoms associated therewith be completely eliminated. In some embodiments, the effect is therapeutic, *i.e.*, without limitation, the effect partially or completely reduces, diminishes, abrogates, abates, alleviates, decreases the intensity of, or cures a disease and/or adverse symptom attributable to the disease. In some embodiments, the effect is preventative, *i.e.*, the effect protects or prevents an occurrence or reoccurrence of a disease or condition. To this end, the presently disclosed methods comprise administering a therapeutically effective amount of a compositions as described herein.

By “uracil glycosylase inhibitor” or “UGI” is meant an agent that inhibits the uracil-excision repair system. Base editors comprising a cytidine deaminase convert cytosine to uracil, which is then converted to thymine through DNA replication or repair. Including an inhibitor of uracil DNA glycosylase (UGI) in the base editor prevents base excision repair which changes the U back to a C. An exemplary UGI comprises an amino acid sequence as follows:

>splP14739IUNGI_BPPB2 Uracil-DNA glycosylase inhibitor

MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSD APEYKPWALVIQDSNGENKIKML (SEQ ID NO: 284).

Ranges provided herein are understood to be shorthand for all of the values within the range. For example, a range of 1 to 50 is understood to include any number, combination of numbers, or sub-range from the group consisting 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50.

The recitation of a listing of chemical groups in any definition of a variable herein includes definitions of that variable as any single group or combination of listed groups. The recitation of an embodiment for a variable or aspect herein includes that embodiment as any single embodiment or in combination with any other embodiments or portions thereof.

All terms are intended to be understood as they would be understood by a person skilled in the art. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the disclosure pertains

In this application, the use of the singular includes the plural unless specifically stated otherwise. It must be noted that, as used in the specification, the singular forms "a," "an" and "the" include plural referents unless the context clearly dictates otherwise. In this application, the use of "or" means "and/or" unless stated otherwise. Furthermore, use of the term "including" as well as other forms, such as "include", "includes," and "included," is not limiting.

As used in this specification and claim(s), the words "comprising" (and any form of comprising, such as "comprise" and "comprises"), "having" (and any form of having, such as "have" and "has"), "including" (and any form of including, such as "includes" and "include") or "containing" (and any form of containing, such as "contains" and "contain") are inclusive or open-ended and do not exclude additional, unrecited elements or method steps. . Any embodiments specified as "comprising" a particular component(s) or element(s) are also contemplated as "consisting of" or "consisting essentially of" the particular component(s) or element(s) in some embodiments. It is contemplated that any embodiment discussed in this specification can be implemented with respect to any method or composition of the present disclosure, and vice versa. Furthermore, compositions of the present disclosure can be used to achieve methods of the present disclosure.

The term "about" or "approximately" means within an acceptable error range for the particular value as determined by one of ordinary skill in the art, which will depend in part on how the value is measured or determined, *i.e.*, the limitations of the measurement system.

Reference in the specification to "some embodiments," "an embodiment," "one embodiment" or "other embodiments" means that a particular feature, structure, or characteristic described in connection with the embodiments is included in at least some embodiments, but not necessarily all embodiments, of the present disclosures.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1A provides a schematic that depicts a mechanism of self-inactivation of a base editor. Two gRNAs direct base editing to occur simultaneously at a target site in the host genome and within the coding region of the base editor. If the base editor being used is an Adenine Base Editor (ABE), the catalytic residues of the deaminase domain (His57 (H57), Glu59 (E59), Cys87 (C87), or Cys90 (C90)) can be inactivated through a single A-to-G edit to install Arg, Gly, Arg, or Arg, respectively at each site. If the base editor being used is a Cytosine Base Editor (CBE), pre-mature stop codons can be installed through a single C-to-T edit at any Arg, Gln, or Trp residue within the editor.

FIG. 1B provides a bar graph depicting base editing activity in HEK293T cells after lipofection of ABE7.10-m and ABE7.10-m variants—containing pre-installed TadA mutations (His57Arg, Glu59Gly, Cys87Arg, or Cys90Arg)—at a genomic site (ABCA4 c.5882G>A) and at the self-inactivation site in TadA (His57, Glu59, Cys87, or Cys90) using two gRNAs.

FIG. 1C provides a schematic depicting the DNA sequence of self-inactivation target sites, His57 and Glu59, within the TadA coding region. The 3' PAM sequence is highlighted gray, and the target nucleotide and its position within the protospacer in each sequence is bold. The nucleotide sequences provided in FIG. 1C in order of occurrence from top-to-bottom correspond to SEQ ID NOs: 458-459. The amino acid sequences provided in FIG. 1C in order of occurrence from top-to-bottom correspond to SEQ ID NOs: 460-461.

FIG. 1D provides a graph depicting the base editing activity in HEK293T cells after lipofection of ABE8.5-m codon variants and two gRNAs targeting a genomic site (ABCA4 c.5882G>A) and the self-inactivation site Glu59 of TadA. Activities of the variants were compared to the activity of ABE8.5-m, which was not provided a self-inactivating gRNA.

FIGs. 1E and 1F provide bar graphs showing the base editing kinetics of ABE8.5-m codon variants and two gRNAs, delivered by AAV2, at a genomic site in ARPE-19 cells and at a TadA catalytic residue of ABE. FIG 1E provides a bar graph depicting a 5-week time course of base editing at a genomic site (ABCA4 c.5882G>A) after AAV2 delivery of ABE8.5-m codon variants and two gRNAs. FIG. 1F provides a bar graph depicting editing at the self-inactivation site— amino acid residue His57 or residue Glu59 of TadA—in the same samples from the 5-week time course.

FIGs. 1G and 1H provide bar graphs showing the base editing kinetics of ABE8.5-m codon variants and two gRNAs, delivered by AAV2, at a genomic site in ARPE-19 cells and at a TadA catalytic residue of ABE, where the self-inactivation edits are assessed by two different methods. FIG. 1G provides a bar graph depicting base editing at a genomic site (ABCA4

c.5882G>A) at two weeks after AAV2 delivery of the ABE8.5-m codon variants and two gRNAs. FIG. 1H provides a bar graph depicting the self-inactivation rates assessed by targeted sequencing of either the DNA from cell lysates or the cDNA generated from mRNA of technical replicate samples in the same experiment.

FIG. 2A provides a diagram of mutations that were made to TadA in order to inactive the editor through alteration of the ABE start codon. Mutations in the DNA and protein sequences are highlighted in black. An alternate, out-of-frame start codon is identified by a gray box. The nucleotide sequences provided in FIG. 2A in order of occurrence from top-to-bottom correspond to SEQ ID NOs: 462-466. The amino acid sequences provided in FIG. 2A in order of occurrence from top-to-bottom correspond to SEQ ID NOs: 467-469.

FIG. 2B provides a bar graph depicting base editing activity at genomic site ABCA4 c.5882G>A in HEK293T cells after lipofection of ABE8.5-m variants containing preinstalled start codon mutations. No self-inactivating gRNA was provided in this experiment.

FIG. 2C provides a diagram showing mutations that were made to ABE8.5-m to incorporate a PAM sequence (NGG) that would allow base editing to occur at Met1 of TadA. The nucleotide sequences provided in FIG. 2C in order of occurrence from top-to-bottom correspond to SEQ ID NOs: 470-476. The amino acid sequences provided in FIG. 2C in order of occurrence from top-to-bottom correspond to SEQ ID NOs: 477-480.

FIG. 2D provides a bar graph depicting base editing activity at genomic site ABCA4 c.5882G>A in HEK293T cells after lipofection of ABE8.5-m variants containing installed PAM sequences in TadA compared to an unmutated control. No self-inactivating gRNA was provided in the experiment.

FIG. 2E provides a bar graph depicting base editing activity in HEK293T cells after lipofection of ABE8.5-m and ABE8.5-m variants at a genomic site (ABCA4 c.5882G>A) and at the self-inactivation site Met1 in TadA using two gRNAs.

FIG. 3A provides a schematic showing a mechanism of self-inactivation of a base editor through the incorporation of an intron in the DNA of an Adenine Base Editor (ABE).

FIGs. 3B and 3C provide bar graphs that depict base editing activity in HEK293T cells after lipofection of ABE variants containing an intron in the coding sequence after or within specific codons (residues) of TadA. FIG. 3B provides a bar graph depicting base editing activity after the incorporation of an intron after residue 87 (NF1, PAX2, EEF1A1, Chimera, SLC50A1, ABCB11, BRSK2, PLXNB3, TMPRSS6, IL32), after residue 62 (Chimera, ABCB11, PLXNB3, IL32), or within residue 23 (Chimera, ABCB11, PLXNB3, IL32) of TadA. FIG. 3B provides a bar graph depicting base editing activity after the incorporation of some additional introns after

residue 87 (ANTXRL, PKHD1L1, PADI1, KRT6C, HMCN2, HMCN2-Salmon, or ENPEP-Gecko) in addition to NF1, PAX2, and EEF1A1. No self-inactivating gRNA was provided in this experiment.

FIG. 3D provides a bar graph depicting base editing activity in HEK293T cells after lipofection of ABE variants containing an intron with a pre-installed edit in either the splice acceptor site or the splice donor site. The introns were located after TadA residue 87 (NF1 acceptor, PAX2 acceptor, EEF1A1 acceptor, Chimera acceptor, ANTXRL acceptor, PKHK1L1 acceptor, PADI1 acceptor, KRT6C acceptor, HMCN2 acceptor, ENPEP-Gecko acceptor, HMCN2-Salmon acceptor, NF1 donor, PAX2 donor, EEF1A1 donor, or Chimera donor). No self-inactivating gRNA was provided in this experiment.

FIG. 3E provides a bar graph depicting base editing activity in HEK293T cells after lipofection of ABE variants containing an intron with a pre-installed edit in the splice acceptor site or the splice donor site. The introns were located after TadA residues 129 (NF1 acceptor, PAX2 acceptor, EEF1A1 acceptor), 59 (NF1 acceptor, PAX2 acceptor, EEF1A1 acceptor), 18 (NF1 acceptor, PAX2 acceptor, EEF1A1 acceptor), and 62 (ABCB11 acceptor), or within residue 23 (ABCB11 donor). No self-inactivating gRNA was provided in this experiment.

FIG. 3F provides a bar graph depicting base editing activity in lipofected HEK293T cells at a genomic site (ABCA4 c.5882G>A) and at the acceptor site of an intron (NF1 or PAX2) placed within TadA after residue 87.

FIG. 3G provides a bar graph depicting base editing activity in lipofected HEK293T cells at a genomic site (ABCA4 c.5882G>A) and at the acceptor site of introns placed within TadA after residue 87 (NF1, PAX2, and EEF1A1) and after residue 62 (ABCB11).

FIG. 3H provides a bar graph depicting base editing activity of ABE8.5-m variants containing introns (NF1, PAX2, or EEF1A1) at various locations within TadA (after residues 87, 129, 59, or 18) with or without preinstalled mutations at the splice acceptor site. No self-inactivating gRNA was provided in this experiment.

FIG. 3I provides a bar graph depicting base editing activity in lipofected HEK293T cells at a genomic site (ABCA4 c.5882G>A) and at the acceptor site of introns (NF1, PAX2, and EEF1A1) placed within TadA after residue 87, 129, 59, and 18.

FIG. 3J provides a bar graph depicting base editing activity in lipofected HEK293T cells at a genomic site (ABCA4 c.5882G>A) and at the acceptor site of introns NF1, PAX2, EEF1A1 ANTXRL, PKHD1L1, PADI1, and ENPEP-Gecko placed within TadA after residue 87.

FIGs. 3K, 3L, and 3M provide bar graphs and a stacked bar graphs depicting base editing activity in HEK293T cells after plasmid lipofection of plasmid DNA encoding a self-inactivating

gRNA, a gRNA targeting the genomic site, and an ABE variant containing an intron in the coding sequence of TadA. FIG. 3K provides a bar graph depicting base editing activity at a genomic site (ABCA4 c.5882G>A) and at the acceptor site of introns NF1 or PAX2 placed within TadA after residue 87, where editing was assessed by targeted sequencing of DNA from cell lysates. FIGs. 3L and 3M provide stacked bar graphs depicting the proportion of splice variants within the ABE8.5-m mRNA assessed by RNA-seq of total mRNA. All analyses in FIGs. 3K, 3L, and 3M were performed on technical replicates in the same experiment.

FIG. 3N provides a bar graphs that depicts base editing activity in ARPE-19 cells at 2 weeks after AAV2 delivery of a self-inactivating gRNA targeting the splice acceptor site, a gRNA targeting the genomic site, and an ABE variant containing an NF1 intron in the coding sequence of TadA at residue 87. Editing was measured at the genomic site by targeted sequencing of genomic DNA and editing at the self-inactivation site is measured both by targeted sequencing of recovered AAV genomes and by RNAseq of the total mRNA from the cells. All measurements were taken on technical replicates in the same experiment.

FIGs.4A-4C provide bar graphs showing a 5-week AAV2 transduction experiment where A>G base conversion was measured at weeks 1, 3, and 5 (x-axis) in ARPE-19 cells, which are a cell line derived from retinal pigment epithelia. FIG.4A provides a bar graph showing the editing at a genomic site (ABCA4 c.5882G>A). FIG.4B provides a bar graph showing editing at a TadA catalytic residue or an intron splice acceptor site, as measured by DNA sequencing. FIG.4C provides a bar graph showing measurements of editing of the same loci via RNA amplicon sequencing. In FIGs. 4A-4C, the term "_scrmbl" indicates that a self-inactivating guide sequence has been scrambled. In FIGs. 4A-4C, the NF1 and PAX2 splice acceptor sites were edited using the guides g235 and g239, respectively (see Table 1C).

FIGs.5A and 5B provide bar graphs showing a 2-week AAV2 transduction experiment in ARPE-19 cells at the indicated days post-transduction (x-axis). The bar graphs each indicate the number of viral genomes added (high, medium, or low) for transducing the cells. The number of viral genomes added for transducing the cells was either High (89k vg/cell), Med (17k vg/cell) or Low (9k vg/cell). FIG.5A provides a bar graph showing the editing rates of a genomic site (ABCA4 c.5882G>A) at days 3, 7 and 14 post transduction for the amounts of virus added. FIG.5B provides a bar graph showing editing at a TadA catalytic residue or an intron splice acceptor site, as measured by DNA sequencing at the indicated time points.

FIGs.6A and 6B provide bar graphs showing a 2-week AAV2 time course transduction experiment in ARPE-19 cells in which editing was measured at days 4, 7, and 14. FIG.6A provides a bar graph of showing editing rates of a genomic site (ABCA4 c.5882G>A), as

measured via next-generation sequencing. FIG.6B provides a bar graph showing editing of a TadA catalytic residue or an intron splice acceptor, as measured via RNA amplicon sequencing.

FIGs.7A and 7B provide bar graphs showing results from a plasmid lipofection in HEK293T cells with editing rates measured at days 2 and 7 post-lipofection. FIG.7A provides a bar graph showing editing of a genomic site (ABCA4 c.5882G>A), as measured via next-generation sequencing. FIG.7B provides a bar graph showing editing of a TadA catalytic residue or an intron splice acceptor site, as measured via RNA amplicon sequencing. In FIGs. 7A and 7B, the term "_scrmbl" indicates that a self-inactivating guide sequence has been scrambled.

FIGs.8A and 8B provide bar graphs showing editing data collected following IV tail vein injection of AAV8 in BALB/c mice. FIG.8A provides a graph showing editing at a genomic site (ABCA4 c.5882G>A), as well as editing of a TadA catalytic residue or an intron splice acceptor site, as measured via both DNA and RNA amplicon sequencing after 1 week post-transduction. FIG.8B provides a graph showing the same outcomes after 4-weeks. In FIGs. 8A and 8B, editing of the genomic site is shown on the left y-axis, and editing of the TadA catalytic residue or intron splice acceptor is shown on the right y-axis. In FIGs. 8A and 8B, the term "_scrmbl" indicates that a self-inactivating guide sequence has been scrambled.

DETAILED DESCRIPTION OF THE INVENTION

The invention features compositions comprising self-inactivating base editors and methods of using such editors. The invention also features polynucleotides that encode bases editors having a heterologous intron for self-inactivation, compositions comprising such polynucleotides, and methods of inactivating a base editor encoded by such polynucleotides.

DNA base editing technology generally utilizes an engineered DNA-binding domain—such as an RNA-guided Cas9 nickase (nCas9)—in a protein fusion with either a cytosine deaminase or an adenine deaminase. Cytosine base editors (CBEs) catalyze the transversion of cytosine to thymine (C > T) through a uracil intermediate, while adenine base editors (ABEs) catalyze the transversion of adenine to guanine (A > G) through a hypoxanthine intermediate (Rees, H. A., & Liu, D. R. (2018). *Base editing: precision chemistry on the genome and transcriptome of living cells*. Nat Rev Genet, 19(12), 770-788). DNA base editing relies on the RNA-guided nCas9 domain to bind at a region of interest in the genome, which displaces the non-target strand of genomic DNA that is extruded from nCas9 as an R-loop, thus exposing these unpaired bases for deamination. The target strand of DNA that is bound to the gRNA is also nicked by the nCas9, which biases cellular DNA mismatch repair toward incorporation of

the mutation installed on the R-loop, rather than resolving to the wildtype base pair of the unedited target strand.

As with all genome-modifying tools, precautions should be taken to protect against undesired off-target edits in the DNA, which are permanent and potentially detrimental (Kim, D., *et al.* (2017). *Genome-wide target specificities of CRISPR RNA-guided programmable deaminases*. Nat Biotechnol, 35(5), 475-480; Liang, P., *et al.* (2019). *Genome-wide profiling of adenine base editor specificity by EndoV-seq. Nature Communications*, 10(1), 67; Zuo, E., *et al.* (2019). *Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos*. Science, 364(6437), 289-292). Situations in which the DNA editor is indefinitely expressed—such as when delivered by AAV (Colella, P., *et al.* (2018). *Emerging Issues in AAV-Mediated In Vivo Gene Therapy. Molecular Therapy* - Methods & Clinical Development, 8, 87-104; Nathwani, A. C., *et al.* (2011). *Long-term Safety and Efficacy Following Systemic Administration of a Self-complementary AAV Vector Encoding Human FIX Pseudotyped With Serotype 5 and 8 Capsid Proteins*. Molecular Therapy, 19(5), 876-885; Nguyen, G. N., *et al.* (2021). *A long-term study of AAV gene therapy in dogs with hemophilia A identifies clonal expansions of transduced liver cells*. Nature Biotechnology, 39(1), 47-55; Niemeyer, G. P., *et al.* (2009). *Long-term correction of inhibitor-prone hemophilia B dogs treated with liver-directed AAV2-mediated factor IX gene therapy*. Blood, 113(4), 797-806) - could be potentially problematic even if off-target activity is very low, as the risk of editing at these sites increases with time of exposure.

In addition, persistence of off-target RNA deamination by base editors, while impermanent, can alter the transcriptomic profile of affected cells (Grunewald, J., *et al.* (2019). *Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors*. Nature, 569(7756), 433-437; Rees, H. A., *et al.* (2019). *Analysis and minimization of cellular RNA editing by DNA adenine base editors*. Sci Adv, 5(5), eaax5717; Zhou, C., *et al.* (2019). *Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis*. Nature, 571(7764), 275-278). Mechanisms for programmed self-inactivation of Cas9 nucleases delivered by AAV have been previously described, wherein the transgene that expresses Cas9 is targeted for double-stranded DNA cleavage in addition to the on-target site within the host genome (Epstein, B. E., & Schaffer, D. V. (2016). *Engineering a Self-Inactivating CRISPR System for AAV Vectors*. Molecular Therapy, 24, S50; Li, A., *et al.* (2019). *A Self-Deleting AAV-CRISPR System for In Vivo Genome Editing*. Mol Ther Methods Clin Dev, 12, 111-122). Thus, the instructions for Cas9 expression are removed from the cell to which it was first delivered.

In order to realize the broadest therapeutic utility of base editing technology, the invention provides methods of attenuating activity and expression of the base editor after delivery methods that may otherwise result in long-term expression. In contrast to CRISPR-Cas nucleases, base editors utilize either a nCas9 or a catalytically inactive "dead" variant (dCas9) in order to avoid the formation of indels that would result from an unaltered Cas9 nuclease (Gaudelli, N. M., *et al.* (2017). *Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage*. Nature, 551(7681), 464-471; Komor, A. C., *et al.* (2016). *Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage*. Nature, 533(7603), 420-424). Self-inactivation of base editor via generating a double stranded break in DNA encoding the is possible but carries several considerations. Nickase Cas9 in base editors can be utilized to generate nicks on both strands encoding the base editor DNA. The sites for each nick may occur at a distance close enough in proximity to favor dissociation of base paired nucleotides, including and up to nicking of each strand to generate a blunt-ended double-stranded DNA break. Additionally, such an approach may require that these nicks be made simultaneously, rather than sequentially, to avoid their re-ligation, and involve at least two additional gRNAs to target the nicks. Base editors that incorporate dCas9 are incapable of using this strategy. Therefore, in one embodiment, the invention provides methods that rely on making single-base edits within the editor DNA to reduce or eliminate further editing activity or expression with a goal of minimizing the potential for both guide-dependent, and guide independent (*e.g.* spurious deamination) activity (Yu, Y., *et al.* (2020). *Cytosine base editors with minimized unguided DNA and RNA off-target events and high on-target activity*. Nature Communications, 11(1), 2052). The invention also provides that any of the four sense codons CAA, CAG, CGA, or TGG—encoding Gln, Arg, and Trp residues— in a CBE can be directly converted to a stop codon though a single, C-to-T base edit (Billon, P., *et al.* (2017). *CRISPR-Mediated Base Editing Enables Efficient Disruption of Eukaryotic Genes through Induction of STOP Codons*. Molecular Cell, 67(6), 1068-1079.e1064). Achieving self-inactivation with an ABE, however, requires alternative approaches because no sense codon can be converted to a nonsense codon through an A-to-G base edit.

Described herein, the invention features compositions and methods to promote self-inactivation of base editors after cellular delivery of their encoding genetic material. The methods of the present invention for the self-inactivation of ABE do not rely on direct conversion of a sense codon to a stop codon, and can be adapted to inactivate CBE using C-to-T single-base edits. These compositions and methods utilize base editing to programmatically

install a single-base mutation into the DNA encoding the editor resulting in ablated DNA editing activity or altered expression.

In one embodiment, the invention is based, at least in part, on the discovery that guide RNAs can direct a base editor to mutate active site residues in the deaminase subunit of the base editor to produce a catalytically dead enzyme and a loss of base editing activity. In another embodiment, the invention is also based, at least in part, on the discovery that targeting the start codon of the base editor for a single-base mutation prevents translation.

In another embodiment, the invention is based, at least in part, on the discovery that, introns can be inserted into a base editor coding sequence (*e.g.*, open reading frame). The introns provide sequences that can be targeted for base editing to disrupt or alter productive splicing of the base editor transcript (*e.g.*, mRNA), resulting in a loss of expression of the base editor (*e.g.*, ABE, CBE). In some embodiments, the base edits are made at the 5' or 3' end of the intron sequence (*e.g.*, in a splice donor or splice acceptor site).

EDITING OF TARGET POLYNUCLEOTIDES

Compositions of the invention are used, for example, to produce gene edits for a defined period of time. Once a desired level of editing has been reached, expression of the base editor is reduced or eliminated by disrupting a splice acceptor or donor site of an intron present in a polynucleotide sequence encoding the base editor.

In general, base editing is carried out to induce therapeutic changes in the genome of a cell of a subject. In some embodiments of the present invention, cells (*in vivo* or *in vitro*) are contacted with two or more guide RNAs and a nucleobase editor polypeptide comprising a nucleic acid programmable DNA binding protein (napDNAbp) (*e.g.*, Cas9), a deaminase (*e.g.*, cytidine deaminase or adenosine deaminase). In some embodiments, cells to be edited are contacted with at least one nucleic acid molecule, wherein the at least one nucleic acid molecule encodes two or more guide RNAs and a nucleobase editor polypeptide, which comprises a nucleic acid programmable DNA binding protein (napDNAbp) (*e.g.*, Cas9) domain, a deaminase (*e.g.*, cytidine deaminase or adenosine deaminase) domain, and where the portion of the nucleic acid molecule encoding the nucleobase editor polypeptide comprises an intron comprising a splice acceptor or splice donor site. In some embodiments, cells to be edited are contacted with at least one nucleic acid molecule, wherein the at least one nucleic acid molecule encodes two or more guide RNAs and a nucleobase editor polypeptide, which comprises a nucleic acid programmable DNA binding protein (napDNAbp) (*e.g.*, Cas9) domain, a cytidine deaminase domain, and where the portion of the nucleic acid molecule encoding the nucleobase editor

polypeptide comprises an intron comprising a splice acceptor or splice donor site. In some embodiments, cells to be edited are contacted with at least one nucleic acid molecule, wherein the at least one nucleic acid molecule encodes two or more guide RNAs and a nucleobase editor polypeptide, which comprises a nucleic acid programmable DNA binding protein (napDNAbp) (*e.g.*, Cas9) domain, an adenosine deaminase domain, and where the portion of the nucleic acid molecule encoding the nucleobase editor polypeptide comprises an intron comprising a splice acceptor or splice donor site. In some embodiments, the at least one nucleic acid molecule encoding two or more guide RNAs and a nucleobase editor polypeptide is delivered to cells by one or more vectors (*e.g.*, AAV vector).

In some embodiments, cells to be edited are contacted with at least one nucleic acid molecule encoding two or more guide RNAs and at least two nucleic acid molecules encoding a split nucleobase editor polypeptide, wherein one nucleic acid molecule encodes an N-terminal fragment of a nucleic acid programmable DNA binding protein (napDNAbp) (*e.g.*, Cas9) domain fused to a split intein-N and a deaminase (*e.g.*, cytidine deaminase or adenosine deaminase) domain, wherein a second nucleic acid molecule encodes a C-terminal fragment of a nucleic acid programmable DNA binding protein (napDNAbp) (*e.g.*, Cas9) domain fused to a split intein-C, and either the first or second nucleic acid molecule includes an intron comprising a splice acceptor or splice donor site. In some embodiments, cells to be edited are contacted with at least one nucleic acid molecule encoding two or more guide RNAs and at least two nucleic acid molecules encoding a split nucleobase editor polypeptide, wherein one nucleic acid molecule encodes an N-terminal fragment of a deaminase (*e.g.*, cytidine deaminase or adenosine deaminase) domain fused to a split intein-N, wherein a second nucleic acid molecule encodes a C-terminal fragment of a deaminase (*e.g.*, cytidine deaminase or adenosine deaminase) domain fused to a split intein-C and a nucleic acid programmable DNA binding protein (napDNAbp) (*e.g.*, Cas9) domain, and either the first or second nucleic acid molecule includes an intron comprising a splice acceptor or splice donor site.

In some embodiments, the at least one nucleic acid molecule encoding two or more guide RNAs and the first and second nucleic acid molecules encoding a split nucleobase editor polypeptide are delivered to cells by one or more vectors (*e.g.*, AAV vector). In some embodiments, the at least one nucleic acid molecule encoding two or more guide RNAs and the first and second nucleic acid molecules encoding a split nucleobase editor polypeptide are each delivered to cells by separate vectors (*e.g.*, AAV vector). In some embodiments, the at least one nucleic acid molecule encoding two or more guide RNAs and the first and second nucleic acid

molecules encoding a split nucleobase editor polypeptide are delivered to cells in the same vector (*e.g.*, AAV vector).

In some embodiments, the nucleic acid molecule encoding the nucleobase editor polypeptide comprises a linker. In some embodiments, the intron is inserted within an open reading frame in the nucleic acid molecule encoding the nucleobase editor polypeptide. In some embodiments, the intron is inserted within the nucleic acid programmable DNA binding protein (napDNAbp) (*e.g.*, Cas9) domain, the deaminase (*e.g.*, cytidine deaminase or adenosine deaminase) domain, or the linker. In some embodiments, the intron is inserted in proximity to a protospacer sequence. In some embodiment, the intron is inserted within about 10 to 30 base pairs of the protospacer sequence. In some embodiments, the protospacer sequence is NGG or NNGRRT. In some embodiments, the intron is between about 10 base pairs to about 500 base pairs in length. In some embodiments, the intron is between about 70 base pairs and 150 base pairs. In some embodiments, the intron is between about 100 base pairs and 200 base pairs.

In some embodiments, the two or more guide RNAs include one or more guide RNAs that direct the a nucleobase editor polypeptide to edit a site in the genome of the cell, and one or more guide RNAs that direct a nucleobase editor polypeptide to edit (*e.g.*, A-to-G or C-to-T base edit) a splice acceptor or a splice donor site present in the intron of the nucleic acid encoding the nucleobase editor polynucleotide. In some embodiments, the gRNA comprises nucleotide analogs. These nucleotide analogs can inhibit degradation of the gRNA from cellular processes.

In various instances, it is advantageous for a spacer sequence to include a 5' and/or a 3' "G" nucleotide. In some cases, for example, any spacer sequence or guide polynucleotide provided herein comprises or further comprises a 5' "G", where, in some embodiments, the 5' "G" is or is not complementary to a target sequence. In some embodiments, the 5' "G" is added to a spacer sequence that does not already contain a 5' "G." For example, it can be advantageous for a guide RNA to include a 5' terminal "G" when the guide RNA is expressed under the control of a U6 promoter or the like because the U6 promoter prefers a "G" at the transcription start site (*see* Cong, L. *et al*. "Multiplex genome engineering using CRISPR/Cas systems. *Science* 339:819-823 (2013) doi: 10.1126/science.1231143). In some cases, a 5' terminal "G" is added to a guide polynucleotide that is to be expressed under the control of a promoter, but is optionally not added to the guide polynucleotide if or when the guide polynucleotide is not expressed under the control of a promoter.

In some embodiments, base editing of the present invention is carried out in a subject *in vivo*. In some embodiments, one or more vectors (*e.g.*, AAV vector) comprising at least one nucleic acid molecule encoding two or more guide RNAs and a nucleobase editor polypeptide,

which comprises a nucleic acid programmable DNA binding protein (napDNAbp) (*e.g.*, Cas9) domain, a deaminase (*e.g.*, cytidine deaminase or adenosine deaminase) domain, and where the portion of the nucleic acid molecule encoding the nucleobase editor polypeptide comprises an intron comprising a splice acceptor or splice donor site, are delivered to a cell in a subject *in vivo*.

In some embodiments, one or more vectors (*e.g.*, AAV vector) comprising at least one nucleic acid molecule encoding one or more guide RNAs, which direct a nucleobase editor polypeptide to edit a site in the genome of the cell, and at least one nucleic acid molecule encoding the nucleobase editor polypeptide, which comprises a nucleic acid programmable DNA binding protein (napDNAbp) (*e.g.*, Cas9) domain, a deaminase (*e.g.*, cytidine deaminase or adenosine deaminase) domain, and an intron comprising a splice acceptor or splice donor site, are delivered to a cell in a subject *in vivo* to edit a site in the genome of the cell. In some embodiments, once a desired level of base editing is achieved in the subject, one or more vectors (*e.g.*, AAV vector) comprising at least one nucleic acid molecule encoding one or more guide RNAs, which target for editing the splice acceptor or splice donor site present in the intron of the nucleic acid molecule encoding the nucleobase editor polynucleotide, is delivered to a cell in a subject *in vivo* to edit (*e.g.*, A-to-G or C-to-T base edit) the splice acceptor or a splice donor site in the intron of the nucleic acid molecule encoding the nucleobase editor polynucleotide, thereby self-inactivating the nucleobase editor polynucleotide to reduce or eliminate base editing activity.

In some embodiments, one or more vectors (*e.g.*, AAV vector) comprising at least one nucleic acid molecule encoding two or more guide RNAs and at least two nucleic acid molecules encoding a split nucleobase editor polypeptide, wherein one nucleic acid molecule encodes an N-terminal fragment of a nucleic acid programmable DNA binding protein (napDNAbp) (*e.g.*, Cas9) domain fused to a split intein-N and a deaminase (*e.g.*, cytidine deaminase or adenosine deaminase) domain, wherein a second nucleic acid molecule encodes a C-terminal fragment of a nucleic acid programmable DNA binding protein (napDNAbp) (*e.g.*, Cas9) domain fused to a split intein-C, and either the first or second nucleic acid molecule includes an intron comprising a splice acceptor or splice donor site, are delivered to a cell in a subject *in vivo*. In some embodiments, one or more vectors (*e.g.*, AAV vector) comprising at least one nucleic acid molecule encoding two or more guide RNAs and at least two nucleic acid molecules encoding a split nucleobase editor polypeptide, wherein one nucleic acid molecule encodes an N-terminal fragment of a deaminase (*e.g.*, cytidine deaminase or adenosine deaminase) domain fused to a split intein-N, wherein a second nucleic acid molecule encodes a C-terminal fragment of a deaminase (*e.g.*, cytidine deaminase or adenosine deaminase) domain fused to a split intein-C

and nucleic acid programmable DNA binding protein (napDNAbp) (*e.g.*, Cas9) domain, and either the first or second nucleic acid molecule includes an intron comprising a splice acceptor or splice donor site, are delivered to a cell in a subject *in vivo*.

In some embodiments, one or more vectors (*e.g.*, AAV vector) comprising at least one nucleic acid molecule encoding one or more guide RNAs, which direct a nucleobase editor polypeptide to edit a site in the genome of the cell, and at least two nucleic acid molecules encoding a split nucleobase editor polypeptide, wherein one nucleic acid molecule encodes an N-terminal fragment of a nucleic acid programmable DNA binding protein (napDNAbp) (*e.g.*, Cas9) domain fused to a split intein-N and a deaminase (*e.g.*, cytidine deaminase or adenosine deaminase) domain, wherein a second nucleic acid molecule encodes a C-terminal fragment of a nucleic acid programmable DNA binding protein (napDNAbp) (*e.g.*, Cas9) domain fused to a split intein-C, and either the first or second nucleic acid molecule includes an intron comprising a splice acceptor or splice donor site, are delivered to a cell in a subject *in vivo* to edit a site in the genome of the cell. In some embodiments, one or more vectors (*e.g.*, AAV vector) comprising at least one nucleic acid molecule encoding one or more guide RNAs, which direct a nucleobase editor polypeptide to edit a site in the genome of the cell, and at least two nucleic acid molecules encoding a split nucleobase editor polypeptide, wherein one nucleic acid molecule encodes an N-terminal fragment of a deaminase (*e.g.*, cytidine deaminase or adenosine deaminase) domain fused to a split intein-N, wherein a second nucleic acid molecule encodes a C-terminal fragment of a deaminase (*e.g.*, cytidine deaminase or adenosine deaminase) domain fused to a split intein-C and a nucleic acid programmable DNA binding protein (napDNAbp) (*e.g.*, Cas9) domain, and either the first or second nucleic acid molecule includes an intron comprising a splice acceptor or splice donor site, are delivered to a cell in a subject *in vivo* to edit a site in the genome of the cell. When the one or more vectors (*e.g.*, AAV vector) are delivered to the cell, the cell will express the N-terminal and the C-terminal fragments of the split nucleobase editor polypeptide, which will join together to form the nucleobase editor polypeptide. In some embodiments, once a desired level of base editing is achieved in the subject, one or more vectors (*e.g.*, AAV vector) comprising at least one nucleic acid molecule encoding one or more guide RNAs, which target for editing the splice acceptor or splice donor site present in the intron of the nucleic acid molecule encoding the nucleobase editor polynucleotide, is delivered to a cell in a subject *in vivo* to edit (*e.g.*, A-to-G or C-to-T base edit) the splice acceptor or a splice donor site present in the nucleic acid molecule encoding the intron of the nucleobase editor polynucleotide, thereby self-inactivating the nucleobase editor polynucleotide to reduce or eliminate base editing activity.

The present invention provides, for example, methods of treating a patient having a disease with a SNP of interest by administering two AAV vectors containing a split intein base editor system as provided herein. In some embodiments, the AAV vectors each encode a portion of a base editor: N-terminal portion fused to an intein-N and a C-terminal portion fused to an intein-C. Encoded in the coding sequence of one or more of the two halves of the base editor is an intron sequence. In some embodiments, a guide RNA targeting the SNP is also included in one of the AAV vectors. In some embodiments, the AAV vectors have a tropism relevant for the diseased cell, tissue, or organ, (*e.g.*, the AAV vector is of a single serotype). When a cell is infected with the two AAV vectors of the base editing system, transcripts encoding the two halves of the base editor are expressed and the intron is spliced out. Upon expressing the polypeptides of the two halves, the base editor is reconstituted by protein splicing in the cell via the split intein tags. In some embodiments, after base editing is performed for a period of time to allow base editing to occur, a third AAV is provided encoding a guide RNA, which in conjunction with the base editor in the cell, targets a donor or acceptor splice site in the intron. When this AAV infects a cell expressing the base editor, it alters the splice site to prevent splicing from occurring. Because a portion of the base editor cannot be appropriately expressed, base editing is inactivated or attenuated in the cell, including at on-target and off-target sites.

The present invention also provides guide RNAs that target the intron of a polynucleotide encoding a self-inactivating base editor. Table 1A provides target intron sequences to be used for gRNAs targeting an intron acceptor or donor site.

Table 1A: Exemplary Target Intron Sequences

Intron Name	Intron SEQ ID NO	Intron Polynucleotide Sequence
NF1	226	GTGAGATCAAATGAAAGTTTCATATAGAAATACAAAACCTAGAGAACTG GCATGTAAGAGAAGCAAAAATTACTTCAGCAAGGCCATGTTAGTAAATT TGCATCTGTTTGTCCACATTAG
PAX2	227	GTAGGTGACAATGCTGCAGCTGCCTAATCTAGGTGGGGGGAACTAAATT GTGGGTGAGCTGCTGAATGGTCTGTAGTCTGAGGCTGGGGTGGGGGGAG ACACAACGTCCCCTCCCTGCAAACCACTGCTATTCTGTCCCTCTCTCTC CTTAG
EEF1A1	228	GTAAGTGGCTTTCAAGACCATTGTTAAAAAGCTCTGGGAATGGCGATTT CATGCTTACATAAATTGGCATGCTTGTGTTTCAG
CBA	229	GTAAGTATCAAGGTTACAAGACAGGTTTAAGGAGACCAATAGAAACTGG GCTTGTCTAGACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGG TCTTACTGACATCCACTTTGCCTTTCTCTCCACAG
SLC50A1	230	GTAAGCACAACTGGGATGGGGTGACAGGGGTGCAAGATTGAAAACTGGC TCCTCTCCTCATAGCAGTTCTTGTGATTTCAG
ABCB11	231	GTAAGAAATGTTATTTTTCAGTAAGTGATTTAGTTATTTTTCCTTTTTT

Intron Name	Intron SEQ ID NO	Intron Polynucleotide Sequence
		CTCATTAAAATTTCTCTAACATCTCCCTCTTCATGTTTTAG
BRSK2	232	GTGAGACCCTAGCCCCCTCAACCCTGCCCTGGCCTCTCCCCAAACCTGC CCCCCCACGCTGACCCCCACACCCGGCCGCCCGCAG
PLXNB3	233	GTGGGTGTCAGAGGCATCGGGGCTGCGGGGTAGGGGGCTGCCCCACCCC TAACGAAGTCTGCTCCTCCAG
TMPRSS6	234	GCAGGGAAGTCCTGCTTCCGTGCCCCACCGGTGCTCAGCTGAGGCTCCC TTGAAAATGCGAGGCTGTTTCCAACTTTGGTCTGTTTCCCTGGCAG
IL32	235	GTGGGGAGTTGGGGTCCCCGAAGGTGAGGACCCTCTGGGGATGAGGGTG CTTCTCTGAGACACTTTCTTTTCCTCACACCTGTTCCTCGCCAGCAG
ANTXRL	236	GTATAGACCCCTTGATCTCCTAACCCTAACCCTAACCCTAACCCTAACC TACAAAATCTTAGAGCATCAGTGGGAGCATCTCACTGTCCAGGCTCAAT ATTTCTTCATTTTCTTGCAG
PKHD1L1	237	GTAATTATGATTAAAGATGGTGATTGTTTATTTTCTTTTATGATTGTCC TTAGTATTATGTAACCTGCAAATTCTATTGCAG
PADI1	238	GTGAGTGACACAAGGTGTTGTCTGGGGAGTGGGGAAGGGGGATGGAAGT GAATCCTGTTGGTGGGGTGGAGAAAGGGCGATCTCAAGAGGGCCACTCT CTCCAG
KRT6C	239	GTAAGCATCTCCACCATCCTTCTGTTTACTCTGATGGGGTCTGCAAAGG GGAGATGATGTATAGGGTTGGGTATCTCTGTAAATGTCAGATGTGAAGT TGATCTTATGACCTTCTGTTCTGCAG
HMCN2	240	GTGAGGGTCTCCCAGGCTGGGCAGGGGGAGGGGGCTGCTGCCTTGATTG CGTCCCAGGACACAGCCCTCCTCCAGCCTGCCCTCGCCTTGCTCATCCC CTCCCCATCTCAGCCCCACCCCCACTAACTCTCTCTCTGCTCTGACTCA G
HMCN2-Salmon	241	GTAATGATTGATTGCAATGTATGATTACAATAATCTCAGTATAAGTTCA GTAATAATAACCTTCCACTGCTGTCCTCTGTGTGCACCCAG
ENPEP-Gecko	242	GTAAATATATACAACAGTTTTTCATTTAAATAAGTGCACGGCACAAATA AGAAAAATATGTCAAAAATGTAACCAATAGTTTTTTTCAAATTTAG

Table 1B provides gRNA sequences for targeting an intron acceptor or donor site. In some embodiments, the gRNA sequence is expressed from a U6 promoter. A lowercase “g” in Table 1B below indicates a 5' mismatch relative to the target sequence.

Table 1B: Exemplary gRNA Sequences. A lowercase "g" represents a 5' mismatch relative to the target sequence.

gRNA ID	Target Site	Target + 3' PAM SEQ ID NO	Target + 3' PAM Sequence	gRNA SEQ ID NO	gRNA Sequence (spacer sequences are in plain text and scaffold sequences are in bold text)
g227	TadA_res87_ABCB11_acceptor	285	GTTTTAGGTCATGTGTGCTGGG	191	gGUUUUAGGUCAUGUGUGCU**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g229	TadA_res87_ABCB11_donor	286	TTTCTTACACAGGGCTCGAAGG	192	gUUUCUUACACAGGGCUCGA**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g231	TadA_res87_EEF1A1_acceptor	287	GTTTCAGGTCATGTGTGCTGGG	193	gGUUUCAGGUCAUGUGUGCU**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g233	TadA_res87_EEF1A1_donor	288	GCCACTTACACAGGGCTCGAAGG	194	GCCACUUACACAGGGCUCGA**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g235	TadA_res87_NF1_acceptor	289	ACATTAGGTCATGTGTGCTGGG	195	gACAUUAGGUCAUGUGUGCU**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g237	TadA_res87_NF1_donor	290	GATCTCACACAGGGCTCGAAGG	196	gGAUCUCACACAGGGCUCGA**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g239	TadA_res87_PAX2_acceptor	291	TCCTTAGGTCATGTGTGCTGGG	197	gUCCUUAGGUCAUGUGUGCU**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g241	TadA_res87_PAX2_donor	292	GTCACCTACACAGGGCTCGAAGG	198	GUCACCUACACAGGGCUCGA**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g243	TadA_res87_SLC50A1_acceptor	293	GATTTCAGGTCATGTGTGCTGGG	199	GAUUUCAGGUCAUGUGUGCU**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA**

gRNA ID	Target Site	Target + 3' PAM SEQ ID NO	Target + 3' PAM Sequence	gRNA SEQ ID NO	gRNA Sequence (spacer sequences are in plain text and scaffold sequences are in bold text)
					ACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
g245	TadA_res87_SLC50A1_donor	294	GTGCTTACACAGGGCTCGAAGG	200	gGUGCUUACACAGGGCUCGA**GUUUUAGAGCUAGA** **AAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA** **ACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g247	TadA_res87_Chimera_acceptor	295	TCCACAGGTCATGTGTGCTGGG	201	gUCCACAGGUCAUGUGUGCU**GUUUUAGAGCUAGA** **AAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA** **ACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g249	TadA_res87_Chimera_donor	296	GATACTTACACAGGGCTCGAAGG	202	GAUACUUACACAGGGCUCGA**GUUUUAGAGCUAGA** **AAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA** **ACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g251	TadA_res62_ABCB11_acceptor	297	TGTTTTAGCTGCGGCAAGGCGG	203	gUGUUUUAGCUGCGGCAAGG**GUUUUAGAGCUAGA** **AAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA** **ACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g253	TadA_res62_ABCB11_donor	298	TTTCTTACAGCCATAATTTCGG	204	gUUUCUUACAGCCAUAAUUU**GUUUUAGAGCUAGA** **AAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA** **ACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g255	TadA_res62_Chimera_acceptor	299	CTCCACAGCTGCGGCAAGGCGG	205	gCUCCACAGCUGCGGCAAGG**GUUUUAGAGCUAGA** **AAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA** **ACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g257	TadA_res62_Chimera_donor	300	GATACTTACAGCCATAATTTCGG	206	GAUACUUACAGCCAUAAUUU**GUUUUAGAGCUAGA** **AAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA** **ACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g259	TadA_res23_ABCB11_acceptor	301	TGTTTTAGGGACGAAAGAGAGG	207	gUGUUUUAGGGACGAAAGAG**GUUUUAGAGCUAGA** **AAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA** **ACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g261	TadA_res23_ABCB11_donor	302	TTACCTGGCTCTCTTAGCCAGG	208	gUUACCUGGCUCUCUUAGCC**GUUUUAGAGCUAGA** **AAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA** **ACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g263	TadA_res23_Chi	303	CTCCACAGGGACGAAAG	209	gCUCCACAGGGACGAAAGAG**GUUUUAGAGCUAGA**

gRNA ID	Target Site	Target + 3' PAM SEQ ID NO	Target + 3' PAM Sequence	gRNA SEQ ID NO	gRNA Sequence (spacer sequences are in plain text and scaffold sequences are in bold text)
	mera_acceptor		AGAGG		**AAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g265	TadA_res23_Chimera_donor	302	TTACCTGGCTCTCTTAG CCAGG	208	gUUACCUGGCUCUCUUAGCCC**GUUUUAGAGCUAGA AAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g267	TadA_res87_ANTXRL_acceptor	304	CTTGCAGGTCATGTGTG CTGGG	210	gCUUGCAGGUCAUGUGUGCU**GUUUUAGAGCUAGA AAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g268	TadA_res87_PKHD1L1_acceptor	305	ATTGCAGGTCATGTGTG CTGGG	211	gAUUGCAGGUCAUGUGUGCU**GUUUUAGAGCUAGA AAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g269	TadA_res87_PADI1_acceptor	306	TCTCCAGGTCATGTGTG CTGGG	212	gUCUCCAGGUCAUGUGUGCU**GUUUUAGAGCUAGA AAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g270	TadA_res87_KRT6C_acceptor	307	TCTGCAGGTCATGTGTG CTGGG	213	gUCUGCAGGUCAUGUGUGCU**GUUUUAGAGCUAGA AAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g271	TadA_res87_HMCN2_acceptor	308	GACTCAGGTCATGTGTG CTGGG	214	gGACUCAGGUCAUGUGUGCU**GUUUUAGAGCUAGA AAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g272	TadA_res87_HMCN2-salmon acceptor	309	GCACCCAGGTCATGTGT GCTGGG	215	GCACCCAGGUCAUGUGUGCU**GUUUUAGAGCUAGA AAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g273	TadA_res87_ENPEP-gecko acceptor	310	AATTTAGGTCATGTGTG CTGGG	216	gAAUUUAGGUCAUGUGUGCU**GUUUUAGAGCUAGA AAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g274	TadA_res129_NF1_acceptor	311	CATTAGGTCGAGATCAC AGAGG	217	gCAUUAGGUCGAGAUCACAG**GUUUUAGAGCUAGA AAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**

gRNA ID	Target Site	Target + 3' PAM SEQ ID NO	Target + 3' PAM Sequence	gRNA SEQ ID NO	gRNA Sequence (spacer sequences are in plain text and scaffold sequences are in bold text)
g275	TadA_res129_PAX2_acceptor	312	CCTTAGGTCGAGATCACAGAGG	218	gCCUUAGGUCGAGAUCACAG**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g276	TadA_res129_EEF1A1_acceptor	313	GTTTCAGGTCGAGATCACAGAGG	219	GUUUCAGGUCGAGAUCACAG**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g277	TadA_res18_NF1_acceptor	314	ACATTAGGCTAAGAGAGCCAGG	220	gACAUUAGGCUAAGAGAGCC**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g278	TadA_res18_PAX2_acceptor	315	TCCTTAGGCTAAGAGAGCCAGG	221	gUCCUUAGGCUAAGAGAGCC**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g279	TadA_res18_EEF1A1_acceptor	316	GTTTCAGGCTAAGAGAGCCAGG	222	gGUUUCAGGCUAAGAGAGCC**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g280	TadA_res59_NF1_acceptor	317	ACATTAGATTATGGCTCTGCGG	223	gACAUUAGAUUAUGGCUCUG**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g281	TadA_res59_PAX_acceptor	318	TCCTTAGATTATGGCTCTGCGG	224	gUCCUUAGAUUAUGGCUCUG**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g282	TadA_res59_EEF1A1_acceptor	319	GTTTCAGATTATGGCTCTGCGG	225	gGUUUCAGAUUAUGGCUCUG**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g316	TadA_start_codon_v5-v8	501	CACCATGAGCGAGGTCGAGNGG	524	gCACCAUGAGCGAGGUCGA**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU**
g318	TadA_start_codon_v9-v10	502	GCCACCATGAGCGAGGTCGAGG	525	gGCCACCAUGAGCGAGGUC**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAA**

gRNA ID	Target Site	Target + 3' PAM SEQ ID NO	Target + 3' PAM Sequence	gRNA SEQ ID NO	gRNA Sequence (spacer sequences are in plain text and scaffold sequences are in bold text)
					CUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
g756	Genomic Site	503	GTGTCGAAGTTCGCCCTGGAGAGG	526	GUGUCGAAGUUCGCCCUGGAG**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU**
gE59G_80	TadA_E59_v80	504	ATGCCGAGATAATGGCCCTCCGG	527	gAUGCCGAGAUAAUGGCCCUC**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU**
gE59G_82	TadA_E59_v82	505	ATGCCGAGATAATGGCCCTTCGG	528	gAUGCCGAGAUAAUGGCCCUU**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU**
gE59G_97	TadA_E59_v97	506	ATGCCGAGATCATGGCACTACGG	529	gAUGCCGAGAUCAUGGCACUA**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU**
gE59G_98	TadA_E59_v98	507	ATGCCGAGATCATGGCACTCCGG	530	gAUGCCGAGAUCAUGGCACUC**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU**
gE59G_99	TadA_E59_v99	508	ATGCCGAGATCATGGCACTGCGG	531	gAUGCCGAGAUCAUGGCACUG**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU**
gE59G_109	TadA_E59_v109	509	ATGCCGAGATCATGGCGCTACGG	532	gAUGCCGAGAUCAUGGCGCUA**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU**

gRNA ID	Target Site	Target + 3' PAM SEQ ID NO	Target + 3' PAM Sequence	gRNA SEQ ID NO	gRNA Sequence (spacer sequences are in plain text and scaffold sequences are in bold text)
gE59G_110	TadA_E59_v110	510	ATGCCGAGATCATGGCG CTCCGG	533	gAUGCCGAGAUCAUGGCGCUC**GUUUUAGAGCUAG AAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUC AACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU U**
gE59G_113	TadA_E59_v113	511	ATGCCGAGATCATGGCG TTACGG	534	gAUGCCGAGAUCAUGGCGUUA**GUUUUAGAGCUAG AAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUC AACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU U**
gE59G_121	TadA_E59_v121	512	ATGCCGAGATTATGGCA CTACGG	535	gAUGCCGAGAUUAUGGCACUA**GUUUUAGAGCUAG AAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUC AACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU U**
gE59G_122	TadA_E59_v122	513	ATGCCGAGATTATGGCA CTCCGG	536	gAUGCCGAGAUUAUGGCACUC**GUUUUAGAGCUAG AAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUC AACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU U**
gE59G_123	TadA_E59_v123	514	ATGCCGAGATTATGGCA CTGCGG	537	gAUGCCGAGAUUAUGGCACUG**GUUUUAGAGCUAG AAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUC AACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU U**
gE59G_124	TadA_E59_v124	515	ATGCCGAGATTATGGCA CTTCGG	538	gAUGCCGAGAUUAUGGCACUU**GUUUUAGAGCUAG AAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUC AACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU U**
gE59G_135	TadA_E59_v135	516	ATGCCGAGATTATGGCG CTGCGG	539	gAUGCCGAGAUUAUGGCGCUG**GUUUUAGAGCUAG AAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUC AACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU U**
gE59G_139	TadA_E59_v139	517	ATGCCGAGATTATGGCT CTACGG	540	gAUGCCGAGAUUAUGGCUCUA**GUUUUAGAGCUAG AAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUC**

gRNA ID	Target Site	Target + 3' PAM SEQ ID NO	Target + 3' PAM Sequence	gRNA SEQ ID NO	gRNA Sequence (spacer sequences are in plain text and scaffold sequences are in bold text)
					AACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU
gE59G_183	TadA_E59_v183	518	ATGCGGAGATCATGGCGCTGCGG	541	gAUGCGGAGAUCAUGGCGCUG**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU**
gE59G_224	TadA_E59_v224	519	ATGCTGAGATAATGGCCCTCCGG	542	gAUGCUGAGAUAAUGGCCCUC**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU**
gH57R	TadA_H57R_v1	520	AACCGCACATGCCGAAATTATGG	543	gAACCGCACAUGCCGAAAUUA**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU**
gE59G_scrmbl	N/A; negative control	N/A	N/A; negative control	544	gGCAGGUGUCGACAUAUCUAU**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU**
gE59G	TadA_E59G_v1	521	ATGCCGAAATTATGGCTCTGCGG	545	gAUGCCGAAAUUAUGGCUCUG**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU**
gC87R	TadA_C87_v1	522	ACACATGACACAGGGCTCGAAGG	546	gACACAUGACACAGGGCUCGA**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU**
gC90R	TadA_C90_v1	523	GCCCCAGCACACATGACACAGGG	547	gGCCCCAGCACACAUGACACA**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU**

In some embodiments, the deaminase domain is a TadA domain. In some embodiments, the intron is inserted within or directly after a codon of TadA. In some embodiments, the intron is inserted within or directly after codon 18, 23, 59, 62, 87, or 129 of TadA. In some embodiments, the intron is inserted directly after codon 87 of TadA.

Table 1C below provides target sequence coordinates for inserting an intron into the TadA open reading frame (*e.g.*, c.100+1 indicates that the first base pair of the intron sequence was directly after the 100th coding nucleotide of TadA). Thus, in some embodiments, the intron sequence is placed directly after a named amino acid position. In other embodiments, the intron sequence is placed directly before a named amino acid position.

Table 1C: Exemplary Target Intron Sequence Coordinates

		Splice Donor Sites		Splice Acceptor Sites	
Intron variant ID	**Codon preceding first base pair of intron insertion position in a TadA reference sequence (e.g., SEQ ID NO: 1)**	**gRNA ID**	**Target Sequence + 3' PAM**	**gRNA ID**	**Target Sequence + 3' PAM**
NF1	Res.18; c.54+1	**g237**	GATCTCACACAGGGCTCGAAGG (SEQ ID NO: 290)	**g277**	ACATTAGGCTAAGAGAGCCAGG (SEQ ID NO: 314)
	Res.59; c.177+1			**g280**	ACATTAGATTATGGCTCTGCGG (SEQ ID NO: 317)
	Res.87; c.261+1			**g235**	ACATTAGGTCATGTGTGCTGGG (SEQ ID NO: 289)
	Res.129; 387+1			**g274**	CATTAGGTCGAGATCACAGAGG (SEQ ID NO: 311)
PAX2	Res.18; c.54+1	**g241**	GTCACCTACACAGGGCTCGAAGG (SEQ ID NO: 292)	**g278**	TCCTTAGGCTAAGAGAGCCAGG (SEQ ID NO: 315)
	Res.59; c.177+1			**g281**	TCCTTAGATTATGGCTCTGCGG (SEQ ID NO: 318)
	Res.87; c.261+1			**g239**	TCCTTAGGTCATGTGTGCTGGG (SEQ ID NO: 291)
	Res.129; 387+1			**g275**	CCTTAGGTCGAGATCACAGAGG (SEQ ID NO: 312)
EEF1A1	Res.18; c.54+1	**g233**	GCCACTTACACAGGGCTCGAAGG (SEQ ID NO: 288)	**g279**	GTTTCAGGCTAAGAGAGCCAGG (SEQ ID NO: 316)
	Res.59; c.177+1			**g282**	GTTTCAGATTATGGCTCTGCGG (SEQ ID NO: 319)
	Res.87; c.261+1			**g231**	GTTTCAGGTCATGTGTGCTGGG (SEQ ID NO: 287)
	Res.129; 387+1			**g276**	GTTTCAGGTCGAGATCACAGAGG (SEQ ID NO: 313)

		Splice Donor Sites		Splice Acceptor Sites	
Intron variant ID	**Codon preceding first base pair of intron insertion position in a TadA reference sequence (e.g., SEQ ID NO: 1)**	**gRNA ID**	**Target Sequence + 3' PAM**	**gRNA ID**	**Target Sequence + 3' PAM**
Chimeric human HBB / mouse Ighg1	Res.23; c.68+1	**g265**	TTACCTGGCTCTCTTAGCCAGG (SEQ ID NO: 302)	**g263**	CTCCACAGGGACGAAAGAGAGG (SEQ ID NO: 303)
	Res.62; c.186+1	**g257**	GATACTTACAGCCATAATTTCGG (SEQ ID NO: 300)	**g255**	CTCCACAGCTGCGGCAAGGCGG (SEQ ID NO: 299)
	Res.87; c.261+1	**g249**	GATACTTACACAGGGCTCGAAGG (SEQ ID NO: 296)	**g247**	TCCACAGGTCATGTGTGCTGGG (SEQ ID NO: 295)
ABCB11	Res.23; c.68+1	**g261**	TTACCTGGCTCTCTTAGCCAGG (SEQ ID NO: 302)	**g259**	TGTTTTAGGGACGAAAGAGAGG (SEQ ID NO: 301)
	Res.62; c.186+1	**g253**	TTTCTTACAGCCATAATTTCGG (SEQ ID NO: 298)	**g251**	TGTTTTAGCTGCGGCAAGGCGG (SEQ ID NO: 297)
	Res.87; c.261+1	**g229**	TTTCTTACACAGGGCTCGAAGG (SEQ ID NO: 286)	**g227**	GTTTTAGGTCATGTGTGCTGGG (SEQ ID NO: 285)
BRSK2	Res.87; c.261+1	**g245**	GTGCTTACACAGGGCTCGAAGG (SEQ ID NO: 294)	**g243**	GATTTCAGGTCATGTGTGCTGGG (SEQ ID NO: 293)
ANTXRL	Res.87; c.261+1			**g267**	CTTGCAGGTCATGTGTGCTGGG (SEQ ID NO: 304)
PKHD1L1	Res.87; c.261+1			**g268**	ATTGCAGGTCATGTGTGCTGGG (SEQ ID NO: 305)
PADI1	Res.87; c.261+1			**g269**	TCTCCAGGTCATGTGTGCTGGG (SEQ ID NO: 306)
KRT6C	Res.87; c.261+1			**g270**	TCTGCAGGTCATGTGTGCTGGG (SEQ ID NO: 307)
HMCN2	Res.87; c.261+1			**g271**	GACTCAGGTCATGTGTGCTGGG (SEQ ID NO: 308)

Intron variant ID	Codon preceding first base pair of intron insertion position in a TadA reference sequence (e.g., SEQ ID NO: 1)	Splice Donor Sites		Splice Acceptor Sites	
		gRNA ID	Target Sequence + 3' PAM	gRNA ID	Target Sequence + 3' PAM
HMCN2-Salmon	Res.87; c.261+1			g272	GCACCCAGGTCATGTGTGCTGGG (SEQ ID NO: 309)
ENPEP-Gecko	Res.87; c.261+1			g273	AATTTAGGTCATGTGTGCTGGG (SEQ ID NO: 310)
BRSK2	Res.87; c.261+1				
PLXNB3	Res.87; c.261+1				
TMPRSS6	Res.87; c.261+1				
IL32	Res.87; c.261+1				

NUCLEOBASE EDITORS

Useful in the methods and compositions described herein are nucleobase editors (*e.g.*, self-inactivating nucleobase editors) that edit, modify or alter a target nucleotide sequence of a polynucleotide. Nucleobase editors described herein typically include a polynucleotide programmable nucleotide binding domain and a nucleobase editing domain (*e.g.*, adenosine deaminase or cytidine deaminase). A polynucleotide programmable nucleotide binding domain, when in conjunction with a bound guide polynucleotide (*e.g.*, gRNA), can specifically bind to a target polynucleotide sequence and thereby localize the base editor to the target nucleic acid sequence desired to be edited. In some embodiments, target polynucleotide sequence is present in an intron (*e.g.*, splice acceptor or splice donor site).

In certain embodiments, the nucleobase editors provided herein comprise one or more features that improve the base editing activity. For example, any of the nucleobase editors provided herein may comprise a Cas9 domain that has reduced nuclease activity. In some embodiments, any of the nucleobase editors provided herein may have a Cas9 domain that does not have nuclease activity (dCas9), or a Cas9 domain that cuts one strand of a duplexed DNA molecule, referred to as a Cas9 nickase (nCas9). Without wishing to be bound by any particular theory, the presence of the catalytic residue (*e.g.*, H840) maintains the activity of the Cas9 to cleave the non-edited (*e.g.*, non-deaminated) strand opposite the targeted nucleobase. Mutation of the catalytic residue (*e.g.*, D10 to A10) prevents cleavage of the edited (*e.g.*, deaminated) strand containing the targeted residue (*e.g.*, A or C). Such Cas9 variants can generate a single-strand DNA break (nick) at a specific location based on the gRNA-defined target sequence, leading to repair of the non-edited strand, ultimately resulting in a nucleobase change on the non-edited strand.

Polynucleotide Programmable Nucleotide Binding Domain

Polynucleotide programmable nucleotide binding domains bind polynucleotides (*e.g.*, RNA, DNA). In some embodiments, an intron is present in an open reading frame encoding a nucleotide programmable nucleotide binding domain of a base editor. A polynucleotide programmable nucleotide binding domain of a base editor can itself comprise one or more domains (*e.g.*, one or more nuclease domains). In some embodiments, the nuclease domain of a polynucleotide programmable nucleotide binding domain can comprise an endonuclease or an exonuclease. An endonuclease can cleave a single strand of a double-stranded nucleic acid or both strands of a double-stranded nucleic acid molecule. In some embodiments, a nuclease

domain of a polynucleotide programmable nucleotide binding domain can cut zero, one, or two strands of a target polynucleotide.

Non-limiting examples of a polynucleotide programmable nucleotide binding domain which can be incorporated into a base editor include a CRISPR protein-derived domain, a restriction nuclease, a meganuclease, TAL nuclease (TALEN), and a zinc finger nuclease (ZFN). In some embodiments, a base editor comprises a polynucleotide programmable nucleotide binding domain comprising a natural or modified protein or portion thereof which via a bound guide nucleic acid is capable of binding to a nucleic acid sequence during CRISPR (*i.e.*, Clustered Regularly Interspaced Short Palindromic Repeats)-mediated modification of a nucleic acid. Such a protein is referred to herein as a "CRISPR protein." Accordingly, disclosed herein is a base editor comprising a polynucleotide programmable nucleotide binding domain comprising all or a portion of a CRISPR protein (*i.e.* a base editor comprising as a domain all or a portion of a CRISPR protein, also referred to as a "CRISPR protein-derived domain" of the base editor). A CRISPR protein-derived domain incorporated into a base editor can be modified compared to a wild-type or natural version of the CRISPR protein. For example, as described below a CRISPR protein-derived domain can comprise one or more mutations, insertions, deletions, rearrangements and/or recombinations relative to a wild-type or natural version of the CRISPR protein.

Cas proteins that can be used herein include class 1 and class 2. Non-limiting examples of Cas proteins include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5d, Cas5t, Cas5h, Cas5a, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 or Csx12), Cas10, Csy1 , Csy2, Csy3, Csy4, Cse1, Cse2, Cse3, Cse4, Cse5e, Csc1, Csc2, Csa5, Csn1, Csn2, Csm1, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx1S, Csf1, Csf2, CsO, Csf4, Csd1, Csd2, Cst1, Cst2, Csh1, Csh2, Csa1, Csa2, Csa3, Csa4, Csa5, Cas12a/Cpf1, Cas12b/C2c1 (*e.g.*, SEQ ID NO: 320), Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, and Cas12j/CasΦ, CARF, DinG, homologues thereof, or modified versions thereof. A CRISPR enzyme can direct cleavage of one or both strands at a target sequence, such as within a target sequence and/or within a complement of a target sequence. For example, a CRISPR enzyme can direct cleavage of one or both strands within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, or more base pairs from the first or last nucleotide of a target sequence.

A vector that encodes a CRISPR enzyme that is mutated to with respect, to a corresponding wild-type enzyme such that the mutated CRISPR enzyme lacks the ability to cleave one or both strands of a target polynucleotide containing a target sequence can be used. A

Cas protein (*e.g.*, Cas9, Cas12) or a Cas domain (*e.g.*, Cas9, Cas12) can refer to a polypeptide or domain with at least or at least about 50%, 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity and/or sequence homology to a wild-type exemplary Cas polypeptide or Cas domain. Cas (*e.g.*, Cas9, Cas12) can refer to the wild-type or a modified form of the Cas protein that can comprise an amino acid change such as a deletion, insertion, substitution, variant, mutation, fusion, chimera, or any combination thereof. In some embodiments, a CRISPR protein-derived domain of a base editor can include all or a portion of Cas9 from *Corynebacterium ulcerans* (NCBI Refs: NC_015683.1, NC_017317.1); *Corynebacterium diphtheria* (NCBI Refs: NC_016782.1, NC_016786.1); *Spiroplasma syrphidicola* (NCBI Ref: NC_021284.1); *Prevotella intermedia* (NCBI Ref: NC_017861.1); *Spiroplasma taiwanense* (NCBI Ref: NC_021846.1); *Streptococcus iniae* (NCBI Ref: NC_021314.1); *Belliella baltica* (NCBI Ref: NC_018010.1); *Psychroflexus torquis* (NCBI Ref: NC_018721.1); *Streptococcus thermophilus* (NCBI Ref: YP_820832.1); *Listeria innocua* (NCBI Ref: NP_472073.1); *Campylobacter jejuni* (NCBI Ref: YP_002344900.1); *Neisseria meningitidis* (NCBI Ref: YP_002342100.1), *Streptococcus pyogenes*, or *Staphylococcus aureus*.

Cas9 nuclease sequences and structures are well known to those of skill in the art (See, *e.g.*, "Complete genome sequence of an Ml strain of Streptococcus pyogenes." Ferretti et al., *Proc. Natl. Acad. Sci. U.S.A.* 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., et al., Nature 471:602-607(2011); and "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., et al., *Science* 337:816-821(2012), the entire contents of each of which are incorporated herein by reference). Cas9 orthologs have been described in various species, including, but not limited to, *S. pyogenes* and S. *thermophilus*. Additional suitable Cas9 nucleases and sequences will be apparent to those of skill in the art based on this disclosure, and such Cas9 nucleases and sequences include Cas9 sequences from the organisms and loci disclosed in Chylinski, Rhun, and Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) *RNA Biology* 10:5, 726-737; the entire contents of which are incorporated herein by reference.

High Fidelity Cas9 Domains

Some aspects of the disclosure provide high fidelity Cas9 domains. High fidelity Cas9 domains are known in the art and described, for example, in Kleinstiver, B.P., *et al.* "High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects." *Nature* 529, 490-495 (2016); and Slaymaker, I.M., *et al.* "Rationally engineered Cas9 nucleases with

improved specificity." *Science* 351, 84-88 (2015); the entire contents of each of which are incorporated herein by reference. An Exemplary high fidelity Cas9 domain is provided in the Sequence Listing as SEQ ID NO: 321. In some embodiments, high fidelity Cas9 domains are engineered Cas9 domains comprising one or more mutations that decrease electrostatic interactions between the Cas9 domain and the sugar-phosphate backbone of a DNA, relative to a corresponding wild-type Cas9 domain. High fidelity Cas9 domains that have decreased electrostatic interactions with the sugar-phosphate backbone of DNA have less off-target effects. In some embodiments, the Cas9 domain (*e.g.*, a wild type Cas9 domain (SEQ ID NOs: 250 and 253)) comprises one or more mutations that decrease the association between the Cas9 domain and the sugar-phosphate backbone of a DNA. In some embodiments, a Cas9 domain comprises one or more mutations that decreases the association between the Cas9 domain and the sugar-phosphate backbone of DNA by at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, or at least 70%.

In some embodiments, any of the Cas9 fusion proteins provided herein comprise one or more of a D10A, N497X, a R661X, a Q695X, and/or a Q926X mutation, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid. .In some embodiments, the high fidelity Cas9 enzyme is SpCas9(K855A), eSpCas9(1.1), SpCas9-HF1, or hyper accurate Cas9 variant (HypaCas9). In some embodiments, the modified Cas9 eSpCas9(1.1) contains alanine substitutions that weaken the interactions between the HNH/RuvC groove and the non-target DNA strand, preventing strand separation and cutting at off-target sites. Similarly, SpCas9-HF1 lowers off-target editing through alanine substitutions that disrupt Cas9's interactions with the DNA phosphate backbone. HypaCas9 contains mutations (SpCas9 N692A/M694A/Q695A/H698A) in the REC3 domain that increase Cas9 proofreading and target discrimination. All three high fidelity enzymes generate less off-target editing than wildtype Cas9.

Cas9 Domains with Reduced Exclusivity

Typically, Cas9 proteins, such as Cas9 from *S. pyogenes* (spCas9), require a "protospacer adjacent motif (PAM)" or PAM-like motif, which is a 2-6 base pair DNA sequence immediately following the DNA sequence targeted by the Cas9 nuclease in the CRISPR bacterial adaptive immune system. The presence of an NGG PAM sequence is required to bind a particular nucleic acid region, where the "N" in "NGG" is adenosine (A), thymidine (T), or cytosine (C), and the G is guanosine. This may limit the ability to edit desired bases within a genome. In some

embodiments, the base editing fusion proteins provided herein may need to be placed at a precise location, for example a region comprising a target base that is upstream of the PAM. See *e.g.*, Komor, A.C., *et al.*, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" *Nature* 533, 420-424 (2016), the entire contents of which are hereby incorporated by reference. Exemplary polypeptide sequences for spCas9 proteins capable of binding a PAM sequence are provided in the Sequence Listing as SEQ ID NOs: 250, 254, and 322-325. Accordingly, in some embodiments, any of the fusion proteins provided herein may contain a Cas9 domain that is capable of binding a nucleotide sequence that does not contain a canonical (*e.g.,* NGG) PAM sequence. Cas9 domains that bind to non-canonical PAM sequences have been described in the art and would be apparent to the skilled artisan. For example, Cas9 domains that bind non-canonical PAM sequences have been described in Kleinstiver, B. P., *et al.*, "Engineered CRISPR-Cas9 nucleases with altered PAM specificities" *Nature* 523, 481-485 (2015); and Kleinstiver, B. P., *et al.*, "Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition" *Nature Biotechnology* 33, 1293-1298 (2015); the entire contents of each are hereby incorporated by reference.

Nickases

In some embodiments, the polynucleotide programmable nucleotide binding domain can comprise a nickase domain. Herein the term "nickase" refers to a polynucleotide programmable nucleotide binding domain comprising a nuclease domain that is capable of cleaving only one strand of the two strands in a duplexed nucleic acid molecule (*e.g.*, DNA). In some embodiments, a nickase can be derived from a fully catalytically active (*e.g.*, natural) form of a polynucleotide programmable nucleotide binding domain by introducing one or more mutations into the active polynucleotide programmable nucleotide binding domain. For example, where a polynucleotide programmable nucleotide binding domain comprises a nickase domain derived from Cas9, the Cas9-derived nickase domain can include a D10A mutation and a histidine at position 840. In such embodiments, the residue H840 retains catalytic activity and can thereby cleave a single strand of the nucleic acid duplex. In another example, a Cas9-derived nickase domain can comprise an H840A mutation, while the amino acid residue at position 10 remains a D. In some embodiments, a nickase can be derived from a fully catalytically active (*e.g.*, natural) form of a polynucleotide programmable nucleotide binding domain by removing all or a portion of a nuclease domain that is not required for the nickase activity. For example, where a polynucleotide programmable nucleotide binding domain comprises a nickase domain derived

from Cas9, the Cas9-derived nickase domain can comprise a deletion of all or a portion of the RuvC domain or the HNH domain.

In some embodiments, wild-type Cas9 corresponds to, or comprises the following amino acid sequence:

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG
NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGN
LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL
SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDH
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVA
YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 250) (single underline: HNH domain; double underline: RuvC domain).

In some embodiments, the strand of a nucleic acid duplex target polynucleotide sequence that is cleaved by a base editor comprising a nickase domain (*e.g.*, Cas9-derived nickase domain, Cas12-derived nickase domain) is the strand that is not edited by the base editor (*i.e.*, the strand that is cleaved by the base editor is opposite to a strand comprising a base to be edited). In other embodiments, a base editor comprising a nickase domain (*e.g.*, Cas9-derived nickase domain,

Cas12-derived nickase domain) can cleave the strand of a DNA molecule which is being targeted for editing. In such embodiments, the non-targeted strand is not cleaved.

In some embodiments, a Cas9 nuclease has an inactive (*e.g.*, an inactivated) DNA cleavage domain, that is, the Cas9 is a nickase, referred to as an "nCas9" protein (for "nickase" Cas9). The Cas9 nickase may be a Cas9 protein that is capable of cleaving only one strand of a duplexed nucleic acid molecule (*e.g.*, a duplexed DNA molecule). In some embodiments the Cas9 nickase cleaves the target strand of a duplexed nucleic acid molecule, meaning that the Cas9 nickase cleaves the strand that is base paired to (complementary to) a gRNA (*e.g.*, an sgRNA) that is bound to the Cas9. In some embodiments, a Cas9 nickase comprises a D10A mutation and has a histidine at position 840. In some embodiments the Cas9 nickase cleaves the non-target, non-base-edited strand of a duplexed nucleic acid molecule, meaning that the Cas9 nickase cleaves the strand that is not base paired to a gRNA (*e.g.*, an sgRNA) that is bound to the Cas9. In some embodiments, a Cas9 nickase comprises an H840A mutation and has an aspartic acid residue at position 10, or a corresponding mutation. In some embodiments the Cas9 nickase comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of the Cas9 nickases provided herein. Additional suitable Cas9 nickases will be apparent to those of skill in the art based on this disclosure and knowledge in the field, and are within the scope of this disclosure.

The amino acid sequence of an exemplary catalytically Cas9 nickase (nCas9) is as follows:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ
TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR

LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKG
KSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 254)

The Cas9 nuclease has two functional endonuclease domains: RuvC and HNH. Cas9 undergoes a conformational change upon target binding that positions the nuclease domains to cleave opposite strands of the target DNA. The end result of Cas9-mediated DNA cleavage is a double-strand break (DSB) within the target DNA (~3-4 nucleotides upstream of the PAM sequence). The resulting DSB is then repaired by one of two general repair pathways: (1) the efficient but error-prone non-homologous end joining (NHEJ) pathway; or (2) the less efficient but high-fidelity homology directed repair (HDR) pathway.

The "efficiency" of non-homologous end joining (NHEJ) and/or homology directed repair (HDR) can be calculated by any convenient method. For example, in some embodiments, efficiency can be expressed in terms of percentage of successful HDR. For example, a surveyor nuclease assay can be used to generate cleavage products and the ratio of products to substrate can be used to calculate the percentage. For example, a surveyor nuclease enzyme can be used that directly cleaves DNA containing a newly integrated restriction sequence as the result of successful HDR. More cleaved substrate indicates a greater percent HDR (a greater efficiency of HDR). As an illustrative example, a fraction (percentage) of HDR can be calculated using the following equation [(cleavage products)/(substrate plus cleavage products)] (*e.g.*, (b+c)/(a+b+c), where "a" is the band intensity of DNA substrate and "b" and "c" are the cleavage products).

In some embodiments, efficiency can be expressed in terms of percentage of successful NHEJ. For example, a T7 endonuclease I assay can be used to generate cleavage products and the ratio of products to substrate can be used to calculate the percentage NHEJ. T7 endonuclease I cleaves mismatched heteroduplex DNA which arises from hybridization of wild-type and mutant DNA strands (NHEJ generates small random insertions or deletions (indels) at the site of the original break). More cleavage indicates a greater percent NHEJ (a greater efficiency of NHEJ). As an illustrative example, a fraction (percentage) of NHEJ can be calculated using the following equation: $(1-(1-(b+c)/(a+b+c))^{1/2})\times 100$, where "a" is the band intensity of DNA

substrate and "b" and "c" are the cleavage products (Ran *et. al.*, *Cell*. 2013 Sep. 12; 154(6):1380-9; and Ran *et al.*, *Nat Protoc*. 2013 Nov.; 8(11): 2281–2308).

The NHEJ repair pathway is the most active repair mechanism, and it frequently causes small nucleotide insertions or deletions (indels) at the DSB site. The randomness of NHEJ-mediated DSB repair has important practical implications, because a population of cells expressing Cas9 and a gRNA or a guide polynucleotide can result in a diverse array of mutations. In most embodiments, NHEJ gives rise to small indels in the target DNA that result in amino acid deletions, insertions, or frameshift mutations leading to premature stop codons within the open reading frame (ORF) of the targeted gene. The ideal end result is a loss-of-function mutation within the targeted gene.

While NHEJ-mediated DSB repair often disrupts the open reading frame of the gene, homology directed repair (HDR) can be used to generate specific nucleotide changes ranging from a single nucleotide change to large insertions like the addition of a fluorophore or tag.

In order to utilize HDR for gene editing, a DNA repair template containing the desired sequence can be delivered into the cell type of interest with the gRNA(s) and Cas9 or Cas9 nickase. The repair template can contain the desired edit as well as additional homologous sequence immediately upstream and downstream of the target (termed left & right homology arms). The length of each homology arm can be dependent on the size of the change being introduced, with larger insertions requiring longer homology arms. The repair template can be a single-stranded oligonucleotide, double-stranded oligonucleotide, or a double-stranded DNA plasmid. The efficiency of HDR is generally low (<10% of modified alleles) even in cells that express Cas9, gRNA and an exogenous repair template. The efficiency of HDR can be enhanced by synchronizing the cells, since HDR takes place during the S and G2 phases of the cell cycle. Chemically or genetically inhibiting genes involved in NHEJ can also increase HDR frequency.

In some embodiments, Cas9 is a modified Cas9. A given gRNA targeting sequence can have additional sites throughout the genome where partial homology exists. These sites are called off-targets and need to be considered when designing a gRNA. In addition to optimizing gRNA design, CRISPR specificity can also be increased through modifications to Cas9. Cas9 generates double-strand breaks (DSBs) through the combined activity of two nuclease domains, RuvC and HNH. Cas9 nickase, a D10A mutant of SpCas9, retains one nuclease domain and generates a DNA nick rather than a DSB. The nickase system can also be combined with HDR-mediated gene editing for specific gene edits.

Catalytically Dead Nucleases

Also provided herein are base editors comprising a polynucleotide programmable nucleotide binding domain which is catalytically dead (*i.e.*, incapable of cleaving a target polynucleotide sequence). Herein the terms "catalytically dead" and "nuclease dead" are used interchangeably to refer to a polynucleotide programmable nucleotide binding domain which has one or more mutations and/or deletions resulting in its inability to cleave a strand of a nucleic acid. In some embodiments, a catalytically dead polynucleotide programmable nucleotide binding domain base editor can lack nuclease activity as a result of specific point mutations in one or more nuclease domains. For example, in the case of a base editor comprising a Cas9 domain, the Cas9 can comprise both a D10A mutation and an H840A mutation. Such mutations inactivate both nuclease domains, thereby resulting in the loss of nuclease activity. In other embodiments, a catalytically dead polynucleotide programmable nucleotide binding domain can comprise one or more deletions of all or a portion of a catalytic domain (*e.g.*, RuvC1 and/or HNH domains). In further embodiments, a catalytically dead polynucleotide programmable nucleotide binding domain comprises a point mutation (*e.g.*, D10A or H840A) as well as a deletion of all or a portion of a nuclease domain. dCas9 domains are known in the art and described, for example, in Qi *et al.*, "Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression." *Cell.* 2013; 152(5):1173-83, the entire contents of which are incorporated herein by reference.

Additional suitable nuclease-inactive dCas9 domains will be apparent to those of skill in the art based on this disclosure and knowledge in the field, and are within the scope of this disclosure. Such additional exemplary suitable nuclease-inactive Cas9 domains include, but are not limited to, D10A/H840A, D10A/D839A/H840A, and D10A/D839A/H840A/N863A mutant domains (See, *e.g.*, Prashant *et al.*, CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology*. 2013; 31(9): 833-838, the entire contents of which are incorporated herein by reference).

In some embodiments, dCas9 corresponds to, or comprises in part or in whole, a Cas9 amino acid sequence having one or more mutations that inactivate the Cas9 nuclease activity. In some embodiments, the nuclease-inactive dCas9 domain comprises a D10X mutation and a H840X mutation of the amino acid sequence set forth herein, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid change. In some embodiments, the nuclease-inactive dCas9 domain comprises a D10A mutation and a H840A mutation of the amino acid sequence set forth herein, or a corresponding mutation in any of the amino acid sequences provided herein. In some embodiments, a nuclease-inactive Cas9 domain

comprises the amino acid sequence set forth in Cloning vector pPlatTET-gRNA2 (Accession No. BAV54124).

In some embodiments, a variant Cas9 protein can cleave the complementary strand of a guide target sequence but has reduced ability to cleave the non-complementary strand of a double stranded guide target sequence. For example, the variant Cas9 protein can have a mutation (amino acid substitution) that reduces the function of the RuvC domain. As a non-limiting example, in some embodiments, a variant Cas9 protein has a D10A (aspartate to alanine at amino acid position 10) and can therefore cleave the complementary strand of a double stranded guide target sequence but has reduced ability to cleave the non-complementary strand of a double stranded guide target sequence (thus resulting in a single strand break (SSB) instead of a double strand break (DSB) when the variant Cas9 protein cleaves a double stranded target nucleic acid) (see, for example, Jinek *et al.*, *Science*. 2012 Aug. 17; 337(6096):816-21).

In some embodiments, a variant Cas9 protein can cleave the non-complementary strand of a double stranded guide target sequence but has reduced ability to cleave the complementary strand of the guide target sequence. For example, the variant Cas9 protein can have a mutation (amino acid substitution) that reduces the function of the HNH domain (RuvC/HNH/RuvC domain motifs). As a non-limiting example, in some embodiments, the variant Cas9 protein has an H840A (histidine to alanine at amino acid position 840) mutation and can therefore cleave the non-complementary strand of the guide target sequence but has reduced ability to cleave the complementary strand of the guide target sequence (thus resulting in a SSB instead of a DSB when the variant Cas9 protein cleaves a double stranded guide target sequence). Such a Cas9 protein has a reduced ability to cleave a guide target sequence (*e.g.*, a single stranded guide target sequence) but retains the ability to bind a guide target sequence (*e.g.*, a single stranded guide target sequence).

As another non-limiting example, in some embodiments, the variant Cas9 protein harbors W476A and W1126A mutations such that the polypeptide has a reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a target DNA (*e.g.*, a single stranded target DNA).

As another non-limiting example, in some embodiments, the variant Cas9 protein harbors P475A, W476A, N477A, D1125A, W1126A, and D1127A mutations such that the polypeptide has a reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a target DNA (*e.g.*, a single stranded target DNA).

As another non-limiting example, in some embodiments, the variant Cas9 protein harbors H840A, W476A, and W1126A, mutations such that the polypeptide has a reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a target DNA (*e.g.*, a single stranded target DNA). As another non-limiting example, in some embodiments, the variant Cas9 protein harbors H840A, D10A, W476A, and W1126A, mutations such that the polypeptide has a reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a target DNA (*e.g.*, a single stranded target DNA). In some embodiments, the variant Cas9 has restored catalytic His residue at position 840 in the Cas9 HNH domain (A840H).

As another non-limiting example, in some embodiments, the variant Cas9 protein harbors, H840A, P475A, W476A, N477A, D1125A, W1126A, and D1127A mutations such that the polypeptide has a reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a target DNA (*e.g.*, a single stranded target DNA). As another non-limiting example, in some embodiments, the variant Cas9 protein harbors D10A, H840A, P475A, W476A, N477A, D1125A, W1126A, and D1127A mutations such that the polypeptide has a reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a target DNA (*e.g.*, a single stranded target DNA). In some embodiments, when a variant Cas9 protein harbors W476A and W1126A mutations or when the variant Cas9 protein harbors P475A, W476A, N477A, D1125A, W1126A, and D1127A mutations, the variant Cas9 protein does not bind efficiently to a PAM sequence. Thus, in some such embodiments, when such a variant Cas9 protein is used in a method of binding, the method does not require a PAM sequence. In other words, in some embodiments, when such a variant Cas9 protein is used in a method of binding, the method can include a guide RNA, but the method can be performed in the absence of a PAM sequence (and the specificity of binding is therefore provided by the targeting segment of the guide RNA). Other residues can be mutated to achieve the above effects (*i.e.*, inactivate one or the other nuclease portions). As non-limiting examples, residues D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, and/or A987 can be altered (*i.e.*, substituted). Also, mutations other than alanine substitutions are suitable.

In some embodiments, a variant Cas9 protein that has reduced catalytic activity (*e.g.*, when a Cas9 protein has a D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, and/or a A987 mutation, *e.g.*, D10A, G12A, G17A, E762A, H840A, N854A, N863A, H982A,

H983A, A984A, and/or D986A), the variant Cas9 protein can still bind to target DNA in a site-specific manner (because it is still guided to a target DNA sequence by a guide RNA) as long as it retains the ability to interact with the guide RNA.

In some embodiments, the variant Cas protein can be spCas9, spCas9-VRQR, spCas9-VRER, xCas9 (sp), saCas9, saCas9-KKH, spCas9-MQKSER, spCas9-LRKIQK, or spCas9-LRVSQL.

In some embodiments, the Cas9 domain is a Cas9 domain from *Staphylococcus aureus* (SaCas9). In some embodiments, the SaCas9 domain is a nuclease active SaCas9, a nuclease inactive SaCas9 (SaCas9d), or a SaCas9 nickase (SaCas9n). In some embodiments, the SaCas9 comprises a N579A mutation, or a corresponding mutation in any of the amino acid sequences provided in the Sequence Listing submitted herewith.

In some embodiments, the SaCas9 domain, the SaCas9d domain, or the SaCas9n domain can bind to a nucleic acid sequence having a non-canonical PAM. In some embodiments, the SaCas9 domain, the SaCas9d domain, or the SaCas9n domain can bind to a nucleic acid sequence having a NNGRRT or a NNGRRV PAM sequence. In some embodiments, the SaCas9 domain comprises one or more of a E781X, a N967X, and a R1014X mutation, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid. In some embodiments, the SaCas9 domain comprises one or more of a E781K, a N967K, and a R1014H mutation, or one or more corresponding mutation in any of the amino acid sequences provided herein. In some embodiments, the SaCas9 domain comprises a E781K, a N967K, or a R1014H mutation, or corresponding mutations in any of the amino acid sequences provided herein.

In some embodiments, one of the Cas9 domains present in the fusion protein may be replaced with a guide nucleotide sequence-programmable DNA-binding protein domain that has no requirements for a PAM sequence. In some embodiments, the Cas9 is an SaCas9. Residue A579 of SaCas9 can be mutated from N579 to yield a SaCas9 nickase. Residues K781, K967, and H1014 can be mutated from E781, N967, and R1014 to yield a SaKKH Cas9.

In some embodiments, a modified SpCas9 including amino acid substitutions D1135M, S1136Q, G1218K, E1219F, A1322R, D1332A, R1335E, and T1337R (SpCas9-MQKFRAER) and having specificity for the altered PAM 5'-NGC-3' was used.

Alternatives to *S. pyogenes* Cas9 can include RNA-guided endonucleases from the Cpf1 family that display cleavage activity in mammalian cells. CRISPR from *Prevotella* and *Francisella 1* (CRISPR/Cpf1) is a DNA-editing technology analogous to the CRISPR/Cas9 system. Cpf1 is an RNA-guided endonuclease of a class II CRISPR/Cas system. This acquired

immune mechanism is found in *Prevotella* and *Francisella* bacteria. Cpf1 genes are associated with the CRISPR locus, coding for an endonuclease that use a guide RNA to find and cleave viral DNA. Cpf1 is a smaller and simpler endonuclease than Cas9, overcoming some of the CRISPR/Cas9 system limitations. Unlike Cas9 nucleases, the result of Cpf1-mediated DNA cleavage is a double-strand break with a short 3′ overhang. Cpf1's staggered cleavage pattern can open up the possibility of directional gene transfer, analogous to traditional restriction enzyme cloning, which can increase the efficiency of gene editing. Like the Cas9 variants and orthologues described above, Cpf1 can also expand the number of sites that can be targeted by CRISPR to AT-rich regions or AT-rich genomes that lack the NGG PAM sites favored by SpCas9. The Cpf1 locus contains a mixed alpha/beta domain, a RuvC-I followed by a helical region, a RuvC-II and a zinc finger-like domain. The Cpf1 protein has a RuvC-like endonuclease domain that is similar to the RuvC domain of Cas9.

Furthermore, Cpf1, unlike Cas9, does not have a HNH endonuclease domain, and the N-terminal of Cpf1 does not have the alpha-helical recognition lobe of Cas9. Cpf1 CRISPR-Cas domain architecture shows that Cpf1 is functionally unique, being classified as Class 2, type V CRISPR system. The Cpf1 loci encode Cas1, Cas2 and Cas4 proteins that are more similar to types I and III than type II systems. Functional Cpf1 does not require the trans-activating CRISPR RNA (tracrRNA), therefore, only CRISPR (crRNA) is required. This benefits genome editing because Cpf1 is not only smaller than Cas9, but also it has a smaller sgRNA molecule (approximately half as many nucleotides as Cas9). The Cpf1-crRNA complex cleaves target DNA or RNA by identification of a protospacer adjacent motif 5'-YTN-3' or 5'-TTN-3' in contrast to the G-rich PAM targeted by Cas9. After identification of PAM, Cpf1 introduces a sticky-end-like DNA double- stranded break having an overhang of 4 or 5 nucleotides.

In some embodiments, the Cas9 is a Cas9 variant having specificity for an altered PAM sequence. In some embodiments, the Additional Cas9 variants and PAM sequences are described in Miller, S.M., *et al.* Continuous evolution of SpCas9 variants compatible with non-G PAMs, Nat. Biotechnol. (2020), the entirety of which is incorporated herein by reference. in some embodiments, a Cas9 variate have no specific PAM requirements. In some embodiments, a Cas9 variant, *e.g.* a SpCas9 variant has specificity for a NRNH PAM, wherein R is A or G and H is A, C, or T. In some embodiments, the SpCas9 variant has specificity for a PAM sequence AAA, TAA, CAA, GAA, TAT, GAT, or CAC. In some embodiments, the SpCas9 variant comprises an amino acid substitution at position 1114, 1134, 1135, 1137, 1139, 1151, 1180, 1188, 1211, 1218, 1219, 1221, 1249, 1256, 1264, 1290, 1318, 1317, 1320, 1321, 1323, 1332, 1333, 1335, 1337, or 1339 or a corresponding position thereof. In some embodiments, the SpCas9 variant comprises

an amino acid substitution at position 1114, 1135, 1218, 1219, 1221, 1249, 1320, 1321, 1323, 1332, 1333, 1335, or 1337 or a corresponding position thereof. In some embodiments, the SpCas9 variant comprises an amino acid substitution at position 1114, 1134, 1135, 1137, 1139, 1151, 1180, 1188, 1211, 1219, 1221, 1256, 1264, 1290, 1318, 1317, 1320, 1323, 1333 or a corresponding position thereof. In some embodiments, the SpCas9 variant comprises an amino acid substitution at position 1114, 1131, 1135, 1150, 1156, 1180, 1191, 1218, 1219, 1221, 1227, 1249, 1253, 1286, 1293, 1320, 1321, 1332, 1335, 1339 or a corresponding position thereof. In some embodiments, the SpCas9 variant comprises an amino acid substitution at position 1114, 1127, 1135, 1180, 1207, 1219, 1234, 1286, 1301, 1332, 1335, 1337, 1338, 1349 or a corresponding position thereof. Exemplary amino acid substitutions and PAM specificity of SpCas9 variants are shown in **Tables 2A-2D**.

Table 2A SpCas9 Variants and PAM specificity

PAM	SpCas9 amino acid position												
	1114	1135	1218	1219	1221	1249	1320	1321	1323	1332	1333	1335	1337
	R	D	G	E	Q	P	A	P	A	D	R	R	T
AAA		N		V	H						G		
AAA		N		V	H						G		
AAA				V							G		
TAA	G	N		V							I		
TAA		N		V							I		A
TAA	G	N		V							I		A
CAA				V							K		
CAA		N		V							K		
CAA		N		V							K		
GAA				V	H		V				K		
GAA		N		V			V				K		
GAA				V	H		V				K		
TAT			S	V	H	S		S				L	
TAT			S	V	H	S		S				L	
TAT			S	V	H	S		S				L	
GAT				V							I		
GAT				V					D			Q	
GAT				V					D			Q	
CAC				V						N		Q	N
CAC		N		V								Q	N
CAC				V						N		Q	N

Table 2B SpCas9 Variants and PAM specificity

SpCas9	SpCas9 amino acid position																			
	1114	1134	1135	1137	1139	1151	1180	1188	1211	1219	1221	1256	1264	1290	1318	1317	1320	1323	1333
	R	F	D	P	V	K	D	K	K	E	Q	Q	H	V	L	N	A	A	R
GAA										V	H						V		K
GAA			N	S						V							V	D	K
GAA			N							V	H		Y				V		K
CAA			N							V	H		Y				V		K
CAA	G		N	S						V	H		Y				V		K
CAA			N					R		V	H						V		K
CAA			N				G		R	V	H		Y				V		K
CAA			N							V	H		Y				V		K
AAA			N				G			V	H	R	Y				V	D	K
CAA	G		N				G			V	H		Y				V	D	K
CAA		L	N				G			V	H		Y			T	V	D	K
TAA	G		N				G			V	H		Y	G	S		V	D	K
TAA	G		N			E	G			V	H		Y		S		V		K
TAA	G		N				G			V	H		Y		S		V	D	K
TAA	G		N				G		R	V	H						V		K
TAA			N				G		R	V	H		Y				V		K
TAA	G		N		A		G			V	H						V		K
TAA	G		N							V	H						V		K

Table 2C SpCas9 Variants and PAM specificity

SpCas9	SpCas9 amino acid position																			
	1114	1131	1135	1150	1156	1180	1191	1218	1219	1221	1227	1249	1253	1286	1293	1320	1321	1332	1335	1339
	R	Y	D	E	K	D	K	G	E	Q	A	P	E	N	A	A	P	D	R	T
SacB.TAT			N				N		V	H						V	S		L	
SacB.TAT			N					S	V	H		S					S	G	L	
AAT			N					S	V	H	V	S		K	T		S	G	L	I
TAT	G		N			G		S	V	H		S	K				S	G	L	
TAT	G		N			G		S	V	H		S					S	G	L	
TAT	G	C	N			G		S	V	H		S					S	G	L	
TAT	G	C	N			G		S	V	H		S					S	G	L	
TAT	G	C	N			G		S	V	H		S					S	G	L	
TAT	G	C	N			G		S	V	H		S					S	G	L	
TAT	G	C	N		E	G		S	V	H		S					S	G	L	
TAT	G	C	N	V		G		S	V	H		S					S	G	L	
TAT		C	N			G		S	V	H		S					S	G	L	
TAT	G	C	N			G		S	V	H		S					S	G	L	

Table 2D SpCas9 Variants and PAM specificity

SpCas9	SpCas9 amino acid position													
	1114	1127	1135	1180	1207	1219	1234	1286	1301	1332	1335	1337	1338	1349
	R	D	D	D	E	E	N	N	P	D	R	T	S	H
SacB.CAC			N			V				N	Q	N		
AAC	G		N			V				N	Q	N		
AAC	G		N			V				N	Q	N		
TAC	G		N			V				N	Q	N		
TAC	G		N			V		H		N	Q	N		
TAC	G		N		G	V	D	H		N	Q	N		
TAC	G		N			V				N	Q	N		

SpCas9	SpCas9 amino acid position													
	1114	1127	1135	1180	1207	1219	1234	1286	1301	1332	1335	1337	1338	1349
	R	D	D	D	E	E	N	N	P	D	R	T	S	H
TAC	G	G	N	E		V		H		N	Q	N		
TAC	G		N			V		H		N	Q	N		
TAC	G		N			V				N	Q	N	T	R

Further exemplary Cas9 (e.g., SaCas9) polypeptides with modified PAM recognition are described in Kleinstiver, et al. "Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition," Nature Biotechnology, 33:1293-1298 (2015) DOI: 10.1038/nbt.3404, the disclosure of which is incorporated herein by reference in its entirety for all purposes. In some embodiments, a Cas9 variant (e.g., a SaCas9 variant) comprising one or more of the alterations E782K, N929R, N968K, and/or R1015H has specificity for, or is associated with increased editing activities relative to a reference polypeptide (e.g., SaCas9) at an NNNRRT or NNHRRT PAM sequence, where N represents any nucleotide, H represents any nucleotide other than G (i.e., "not G"), and R represents a purine. In embodiments, the Cas9 variant (e.g., a SaCas9 variant) comprises the alterations E782K, N968K, and R1015H or the alterations E782K, K929R, and R1015H.

In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) is a single effector of a microbial CRISPR-Cas system. Single effectors of microbial CRISPR-Cas systems include, without limitation, Cas9, Cpf1, Cas12b/C2c1, and Cas12c/C2c3. Typically, microbial CRISPR-Cas systems are divided into Class 1 and Class 2 systems. Class 1 systems have multisubunit effector complexes, while Class 2 systems have a single protein effector. For example, Cas9 and Cpf1 are Class 2 effectors. In addition to Cas9 and Cpf1, three distinct Class 2 CRISPR-Cas systems (Cas12b/C2c1, and Cas12c/C2c3) have been described by Shmakov *et al.*, "Discovery and Functional Characterization of Diverse Class 2 CRISPR Cas Systems", *Mol. Cell,* 2015 Nov. 5; 60(3): 385-397, the entire contents of which is hereby incorporated by reference. Effectors of two of the systems, Cas12b/C2c1, and Cas12c/C2c3, contain RuvC-like endonuclease domains related to Cpf1. A third system contains an effector with two predicated HEPN RNase domains. Production of mature CRISPR RNA is tracrRNA-independent, unlike production of CRISPR RNA by Cas12b/C2c1. Cas12b/C2c1 depends on both CRISPR RNA and tracrRNA for DNA cleavage.

In some embodiments, the napDNAbp is a circular permutant (*e.g.*, SEQ ID NO: 326).

The crystal structure of *Alicyclobaccillus acidoterrastris* Cas12b/C2c1 (AacC2c1) has been reported in complex with a chimeric single-molecule guide RNA (sgRNA). See *e.g.*, Liu *et al.*, "C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism", *Mol. Cell,* 2017 Jan. 19; 65(2):310-322, the entire contents of which are hereby incorporated by reference. The crystal structure has also been reported in *Alicyclobacillus acidoterrestris* C2c1 bound to target DNAs as ternary complexes. See *e.g.*, Yang *et al.*,

"PAM-dependent Target DNA Recognition and Cleavage by C2C1 CRISPR-Cas endonuclease", *Cell,* 2016 Dec. 15; 167(7):1814-1828, the entire contents of which are hereby incorporated by reference. Catalytically competent conformations of AacC2c1, both with target and non-target DNA strands, have been captured independently positioned within a single RuvC catalytic pocket, with Cas12b/C2c1-mediated cleavage resulting in a staggered seven-nucleotide break of target DNA. Structural comparisons between Cas12b/C2c1 ternary complexes and previously identified Cas9 and Cpf1 counterparts demonstrate the diversity of mechanisms used by CRISPR-Cas9 systems.

In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) of any of the fusion proteins provided herein may be a Cas12b/C2c1, or a Cas12c/C2c3 protein. In some embodiments, the napDNAbp is a Cas12b/C2c1 protein. In some embodiments, the napDNAbp is a Cas12c/C2c3 protein. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at ease 99.5% identical to a naturally-occurring Cas12b/C2c1 or Cas12c/C2c3 protein. In some embodiments, the napDNAbp is a naturally-occurring Cas12b/C2c1 or Cas12c/C2c3 protein. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at ease 99.5% identical to any one of the napDNAbp sequences provided herein. It should be appreciated that Cas12b/C2c1 or Cas12c/C2c3 from other bacterial species may also be used in accordance with the present disclosure.

In some embodiments, a napDNAbp refers to Cas12c. In some embodiments, the Cas12c protein is a Cas12c1 (SEQ ID NO: 327) or a variant of Cas12c1. In some embodiments, the Cas12 protein is a Cas12c2 (SEQ ID NO: 328) or a variant of Cas12c2. In some embodiments, the Cas12 protein is a Cas12c protein from *Oleiphilus* sp. HI0009 (*i.e.*, OspCas12c; SEQ ID NO: 329) or a variant of OspCas12c. These Cas12c molecules have been described in Yan *et al.*, "Functionally Diverse Type V CRISPR-Cas Systems," Science, 2019 Jan. 4; 363: 88-91; the entire contents of which is hereby incorporated by reference. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a naturally-occurring Cas12c1, Cas12c2, or OspCas12c protein. In some embodiments, the napDNAbp is a naturally-occurring Cas12c1, Cas12c2, or OspCas12c protein. In some embodiments, the

napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at ease 99.5% identical to any Cas12c1, Cas12c2, or OspCas12c protein described herein. It should be appreciated that Cas12c1, Cas12c2, or OspCas12c from other bacterial species may also be used in accordance with the present disclosure.

In some embodiments, a napDNAbp refers to Cas12g, Cas12h, or Cas12i, which have been described in, for example, Yan *et al.*, "Functionally Diverse Type V CRISPR-Cas Systems," Science, 2019 Jan. 4; 363: 88-91; the entire contents of each is hereby incorporated by reference. Exemplary Cas12g, Cas12h, and Cas12i polypeptide sequences are provided in the Sequence Listing as SEQ ID NOs: 330-333. By aggregating more than 10 terabytes of sequence data, new classifications of Type V Cas proteins were identified that showed weak similarity to previously characterized Class V protein, including Cas12g, Cas12h, and Cas12i. In some embodiments, the Cas12 protein is a Cas12g or a variant of Cas12g. In some embodiments, the Cas12 protein is a Cas12h or a variant of Cas12h. In some embodiments, the Cas12 protein is a Cas12i or a variant of Cas12i. It should be appreciated that other RNA-guided DNA binding proteins may be used as a napDNAbp, and are within the scope of this disclosure. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a naturally-occurring Cas12g, Cas12h, or Cas12i protein. In some embodiments, the napDNAbp is a naturally-occurring Cas12g, Cas12h, or Cas12i protein. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at ease 99.5% identical to any Cas12g, Cas12h, or Cas12i protein described herein. It should be appreciated that Cas12g, Cas12h, or Cas12i from other bacterial species may also be used in accordance with the present disclosure. In some embodiments, the Cas12i is a Cas12i1 or a Cas12i2.

In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) of any of the fusion proteins provided herein may be a Cas12j/CasΦ protein. Cas12j/CasΦ is described in Pausch *et al.*, "CRISPR-CasΦ from huge phages is a hypercompact genome editor," *Science*, 17 July 2020, Vol. 369, Issue 6501, pp. 333-337, which is incorporated herein by reference in its entirety. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%,

at least 99%, or at ease 99.5% identical to a naturally-occurring Cas12j/CasΦ protein. In some embodiments, the napDNAbp is a naturally-occurring Cas12j/CasΦ protein. In some embodiments, the napDNAbp is a nuclease inactive ("dead") Cas12j/CasΦ protein. It should be appreciated that Cas12j/CasΦ from other species may also be used in accordance with the present disclosure.

Fusion Proteins with Internal Insertions

Provided herein are fusion proteins comprising a heterologous polypeptide fused to a nucleic acid programmable nucleic acid binding protein, for example, a napDNAbp. As detailed below, this disclosure provides polynucleotides encoding fusion proteins that feature heterologous polypeptides, where the polynucleotide includes an intron in an open reading frame that encodes all or a portion of a heterologous domain of a fusion protein. A heterologous polypeptide can be a polypeptide that is not found in the native or wild-type napDNAbp polypeptide sequence. The heterologous polypeptide can be fused to the napDNAbp at a C-terminal end of the napDNAbp, an N-terminal end of the napDNAbp, or inserted at an internal location of the napDNAbp. In some embodiments, the heterologous polypeptide is a deaminase (*e.g.*, cytidine of adenosine deaminase) or a functional fragment thereof. For example, a fusion protein can comprise a deaminase flanked by an N- terminal fragment and a C-terminal fragment of a Cas9 or Cas12 (*e.g.*, Cas12b/C2c1), polypeptide. In some embodiments, the cytidine deaminase is an APOBEC deaminase (*e.g.*, APOBEC1). In some embodiments, the adenosine deaminase is a TadA (*e.g.*, TadA*7.10 or TadA*8). In some embodiments, the TadA is a TadA*8 or a TadA*9. TadA sequences (*e.g.*, TadA7.10 or TadA*8) as described herein are suitable deaminases for the above-described fusion proteins.

In some embodiments, the fusion protein comprises the structure:

NH2-[N-terminal fragment of a napDNAbp]-[deaminase]-[C-terminal fragment of a napDNAbp]-COOH;

NH2-[N-terminal fragment of a Cas9]-[adenosine deaminase]-[C-terminal fragment of a Cas9]-COOH;

NH2-[N-terminal fragment of a Cas12]-[adenosine deaminase]-[C-terminal fragment of a Cas12]-COOH;

NH2-[N-terminal fragment of a Cas9]-[cytidine deaminase]-[C-terminal fragment of a Cas9]-COOH;

NH2-[N-terminal fragment of a Cas12]-[cytidine deaminase]-[C-terminal fragment of a Cas12]-COOH;

wherein each instance of "]-[" is an optional linker.

The deaminase can be a circular permutant deaminase. For example, the deaminase can be a circular permutant adenosine deaminase. In some embodiments, the deaminase is a circular permutant TadA, circularly permutated at amino acid residue 116, 136, or 65 as numbered in the TadA reference sequence.

The fusion protein can comprise more than one deaminase. The fusion protein can comprise, for example, 1, 2, 3, 4, 5 or more deaminases. In some embodiments, the fusion protein comprises one or two deaminase. The two or more deaminases in a fusion protein can be an adenosine deaminase, a cytidine deaminase, or a combination thereof. The two or more deaminases can be homodimers or heterodimers. The two or more deaminases can be inserted in tandem in the napDNAbp. In some embodiments, the two or more deaminases may not be in tandem in the napDNAbp.

In some embodiments, the napDNAbp in the fusion protein is a Cas9 polypeptide or a fragment thereof. The Cas9 polypeptide can be a variant Cas9 polypeptide. In some embodiments, the Cas9 polypeptide is a Cas9 nickase (nCas9) polypeptide or a fragment thereof. In some embodiments, the Cas9 polypeptide is a nuclease dead Cas9 (dCas9) polypeptide or a fragment thereof. The Cas9 polypeptide in a fusion protein can be a full-length Cas9 polypeptide. In some cases, the Cas9 polypeptide in a fusion protein may not be a full length Cas9 polypeptide. The Cas9 polypeptide can be truncated, for example, at a N-terminal or C-terminal end relative to a naturally-occurring Cas9 protein. The Cas9 polypeptide can be a circularly permuted Cas9 protein. The Cas9 polypeptide can be a fragment, a portion, or a domain of a Cas9 polypeptide, that is still capable of binding the target polynucleotide and a guide nucleic acid sequence.

In some embodiments, the Cas9 polypeptide is a Streptococcus pyogenes Cas9 (SpCas9), Staphylococcus aureus Cas9 (SaCas9), Streptococcus thermophilus 1 Cas9 (St1Cas9), or fragments or variants of any of the Cas9 polypeptides described herein.

In some embodiments, the fusion protein comprises an adenosine deaminase domain and a cytidine deaminase domain inserted within a Cas9. In some embodiments, an adenosine deaminase is fused within a Cas9 and a cytidine deaminase is fused to the C-terminus. In some embodiments, an adenosine deaminase is fused within Cas9 and a cytidine deaminase fused to the N-terminus. In some embodiments, a cytidine deaminase is fused within Cas9 and an adenosine deaminase is fused to the C-terminus. In some embodiments, a cytidine deaminase is fused within Cas9 and an adenosine deaminase fused to the N-terminus.

Exemplary structures of a fusion protein with an adenosine deaminase and a cytidine deaminase and a Cas9 are provided as follows:

NH2-[Cas9(adenosine deaminase)]-[cytidine deaminase]-COOH;

NH2-[cytidine deaminase]-[Cas9(adenosine deaminase)]-COOH;

NH2-[Cas9(cytidine deaminase)]-[adenosine deaminase]-COOH; or

NH2-[adenosine deaminase]-[Cas9(cytidine deaminase)]-COOH.

In some embodiments, the "-" used in the general architecture above indicates the presence of an optional linker.

In various embodiments, the catalytic domain has DNA modifying activity (*e.g.*, deaminase activity), such as adenosine deaminase activity. In some embodiments, the adenosine deaminase is a TadA (*e.g.*, TadA*7.10). In some embodiments, the TadA is a TadA*8. In some embodiments, a TadA*8 is fused within Cas9 and a cytidine deaminase is fused to the C-terminus. In some embodiments, a TadA*8 is fused within Cas9 and a cytidine deaminase fused to the N-terminus. In some embodiments, a cytidine deaminase is fused within Cas9 and a TadA*8 is fused to the C-terminus. In some embodiments, a cytidine deaminase is fused within Cas9 and a TadA*8 fused to the N-terminus. Exemplary structures of a fusion protein with a TadA*8 and a cytidine deaminase and a Cas9 are provided as follows:

NH2-[Cas9(TadA*8)]-[cytidine deaminase]-COOH;

NH2-[cytidine deaminase]-[Cas9(TadA*8)]-COOH;

NH2-[Cas9(cytidine deaminase)]-[TadA*8]-COOH; or

NH2-[TadA*8]-[Cas9(cytidine deaminase)]-COOH.

In some embodiments, the "-" used in the general architecture above indicates the presence of an optional linker.

The heterologous polypeptide (*e.g.*, deaminase) can be inserted in the napDNAbp (*e.g.*, Cas9 or Cas12 (*e.g.*, Cas12b/C2c1)) at a suitable location, for example, such that the napDNAbp retains its ability to bind the target polynucleotide and a guide nucleic acid. A deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) can be inserted into a napDNAbp without compromising function of the deaminase (*e.g.*, base editing activity) or the napDNAbp (*e.g.*, ability to bind to target nucleic acid and guide nucleic acid). A deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) can be inserted in the napDNAbp at, for example, a disordered region or a region comprising a high temperature factor or B-factor as shown by crystallographic studies. Regions of a protein that are less ordered, disordered, or

unstructured, for example solvent exposed regions and loops, can be used for insertion without compromising structure or function. A deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase)can be inserted in the napDNAbp in a flexible loop region or a solvent-exposed region. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted in a flexible loop of the Cas9 or the Cas12b/C2c1 polypeptide.

In some embodiments, the insertion location of a deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is determined by B-factor analysis of the crystal structure of Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted in regions of the Cas9 polypeptide comprising higher than average B-factors (*e.g.*, higher B factors compared to the total protein or the protein domain comprising the disordered region). B-factor or temperature factor can indicate the fluctuation of atoms from their average position (for example, as a result of temperature-dependent atomic vibrations or static disorder in a crystal lattice). A high B-factor (*e.g.*, higher than average B-factor) for backbone atoms can be indicative of a region with relatively high local mobility. Such a region can be used for inserting a deaminase without compromising structure or function. A deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) can be inserted at a location with a residue having a Cα atom with a B-factor that is 50%, 60%, 70%, 80%, 90%, 100%, 110%, 120%, 130%, 140%, 150%, 160%, 170%, 180%, 190%, 200%, or greater than 200% more than the average B-factor for the total protein. A deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) can be inserted at a location with a residue having a Cα atom with a B-factor that is 50%, 60%, 70%, 80%, 90%, 100%, 110%, 120%, 130%, 140%, 150%, 160%, 170%, 180%, 190%, 200% or greater than 200% more than the average B-factor for a Cas9 protein domain comprising the residue. Cas9 polypeptide positions comprising a higher than average B-factor can include, for example, residues 768, 792, 1052, 1015, 1022, 1026, 1029, 1067, 1040, 1054, 1068, 1246, 1247, and 1248 as numbered in the above Cas9 reference sequence. Cas9 polypeptide regions comprising a higher than average B-factor can include, for example, residues 792-872, 792-906, and 2-791 as numbered in the above Cas9 reference sequence.

A heterologous polypeptide (*e.g.*, deaminase) can be inserted in the napDNAbp at an amino acid residue selected from the group consisting of: 768, 791, 792, 1015, 1016, 1022,

1023, 1026, 1029, 1040, 1052, 1054, 1067, 1068, 1069, 1246, 1247, and 1248 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the heterologous polypeptide is inserted between amino acid positions 768-769, 791-792, 792-793, 1015-1016, 1022-1023, 1026-1027, 1029-1030, 1040-1041, 1052-1053, 1054-1055, 1067-1068, 1068-1069, 1247-1248, or 1248-1249 as numbered in the above Cas9 reference sequence or corresponding amino acid positions thereof. In some embodiments, the heterologous polypeptide is inserted between amino acid positions 769-770, 792-793, 793-794, 1016-1017, 1023-1024, 1027-1028, 1030-1031, 1041-1042, 1053-1054, 1055-1056, 1068-1069, 1069-1070, 1248-1249, or 1249-1250 as numbered in the above Cas9 reference sequence or corresponding amino acid positions thereof. In some embodiments, the heterologous polypeptide replaces an amino acid residue selected from the group consisting of: 768, 791, 792, 1015, 1016, 1022, 1023, 1026, 1029, 1040, 1052, 1054, 1067, 1068, 1069, 1246, 1247, and 1248 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. It should be understood that the reference to the above Cas9 reference sequence with respect to insertion positions is for illustrative purposes. The insertions as discussed herein are not limited to the Cas9 polypeptide sequence of the above Cas9 reference sequence, but include insertion at corresponding locations in variant Cas9 polypeptides, for example a Cas9 nickase (nCas9), nuclease dead Cas9 (dCas9), a Cas9 variant lacking a nuclease domain, a truncated Cas9, or a Cas9 domain lacking partial or complete HNH domain.

A heterologous polypeptide (*e.g.*, deaminase) can be inserted in the napDNAbp at an amino acid residue selected from the group consisting of: 768, 792, 1022, 1026, 1040, 1068, and 1247 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the heterologous polypeptide is inserted between amino acid positions 768-769, 792-793, 1022-1023, 1026-1027, 1029-1030, 1040-1041, 1068-1069, or 1247-1248 as numbered in the above Cas9 reference sequence or corresponding amino acid positions thereof. In some embodiments, the heterologous polypeptide is inserted between amino acid positions 769-770, 793-794, 1023-1024, 1027-1028, 1030-1031, 1041-1042, 1069-1070, or 1248-1249 as numbered in the above Cas9 reference sequence or corresponding amino acid positions thereof. In some embodiments, the heterologous polypeptide replaces an amino acid residue selected from the group consisting of: 768, 792, 1022, 1026, 1040, 1068, and 1247 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

A heterologous polypeptide (*e.g.*, deaminase) can be inserted in the napDNAbp at an amino acid residue as described herein, or a corresponding amino acid residue in another Cas9 polypeptide. In an embodiment, a heterologous polypeptide (*e.g.*, deaminase) can be inserted in the napDNAbp at an amino acid residue selected from the group consisting of: 1002, 1003, 1025, 1052-1056, 1242-1247, 1061-1077, 943-947, 686-691, 569-578, 530-539, and 1060-1077 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. The deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) can be inserted at the N-terminus or the C-terminus of the residue or replace the residue. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of the residue.

In some embodiments, an adenosine deaminase (*e.g.*, TadA) is inserted at an amino acid residue selected from the group consisting of: 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, and 1246 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, an adenosine deaminase (*e.g.*, TadA) is inserted in place of residues 792-872, 792-906, or 2-791 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the adenosine deaminase is inserted at the N-terminus of an amino acid selected from the group consisting of: 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, and 1246 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the adenosine deaminase is inserted at the C-terminus of an amino acid selected from the group consisting of: 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, and 1246 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the adenosine deaminase is inserted to replace an amino acid selected from the group consisting of: 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, and 1246 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, a cytidine deaminase (*e.g.*, APOBEC1) is inserted at an amino acid residue selected from the group consisting of: 1016, 1023, 1029, 1040, 1069, and 1247 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the cytidine deaminase is inserted at the N-terminus of an amino acid selected from the group consisting of: 1016, 1023, 1029, 1040,

1069, and 1247 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the cytidine deaminase is inserted at the C-terminus of an amino acid selected from the group consisting of: 1016, 1023, 1029, 1040, 1069, and 1247 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the cytidine deaminase is inserted to replace an amino acid selected from the group consisting of: 1016, 1023, 1029, 1040, 1069, and 1247 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid residue 768 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid residue 768 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of amino acid residue 768 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted to replace amino acid residue 768 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid residue 791 or is inserted at amino acid residue 792, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid residue 791 or is inserted at the N-terminus of amino acid 792, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of amino acid 791 or is inserted at the N-terminus of amino acid 792, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the

deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted to replace amino acid 791, or is inserted to replace amino acid 792, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid residue 1016 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid residue 1016 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of amino acid residue 1016 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted to replace amino acid residue 1016 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid residue 1022, or is inserted at amino acid residue 1023, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid residue 1022 or is inserted at the N-terminus of amino acid residue 1023, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of amino acid residue 1022 or is inserted at the C-terminus of amino acid residue 1023, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted to replace amino acid residue 1022, or is inserted to replace amino acid residue 1023, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid residue 1026, or is inserted at amino acid residue 1029, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid residue 1026 or is inserted at the N-terminus of amino acid residue 1029, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of amino acid residue 1026 or is inserted at the C-terminus of amino acid residue 1029, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted to replace amino acid residue 1026, or is inserted to replace amino acid residue 1029, as numbered in the above Cas9 reference sequence, or corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid residue 1040 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid residue 1040 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of amino acid residue 1040 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted to replace amino acid residue 1040 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid residue 1052, or is inserted at amino acid residue 1054, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the

deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid residue 1052 or is inserted at the N-terminus of amino acid residue 1054, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of amino acid residue 1052 or is inserted at the C-terminus of amino acid residue 1054, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted to replace amino acid residue 1052, or is inserted to replace amino acid residue 1054, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid residue 1067, or is inserted at amino acid residue 1068, or is inserted at amino acid residue 1069, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid residue 1067 or is inserted at the N-terminus of amino acid residue 1068 or is inserted at the N-terminus of amino acid residue 1069, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of amino acid residue 1067 or is inserted at the C-terminus of amino acid residue 1068 or is inserted at the C-terminus of amino acid residue 1069, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted to replace amino acid residue 1067, or is inserted to replace amino acid residue 1068, or is inserted to replace amino acid residue 1069, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid residue 1246, or is inserted at amino acid residue 1247, or is inserted at amino acid residue 1248, as numbered in

the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid residue 1246 or is inserted at the N-terminus of amino acid residue 1247 or is inserted at the N-terminus of amino acid residue 1248, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of amino acid residue 1246 or is inserted at the C-terminus of amino acid residue 1247 or is inserted at the C-terminus of amino acid residue 1248, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted to replace amino acid residue 1246, or is inserted to replace amino acid residue 1247, or is inserted to replace amino acid residue 1248, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, a heterologous polypeptide (*e.g.*, deaminase) is inserted in a flexible loop of a Cas9 polypeptide. The flexible loop portions can be selected from the group consisting of 530-537, 569-570, 686-691, 943-947, 1002-1025, 1052-1077, 1232-1247, or 1298-1300 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. The flexible loop portions can be selected from the group consisting of: 1-529, 538-568, 580-685, 692-942, 948-1001, 1026-1051, 1078-1231, or 1248-1297 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

A heterologous polypeptide (*e.g.*, adenine deaminase) can be inserted into a Cas9 polypeptide region corresponding to amino acid residues: 1017-1069, 1242-1247, 1052-1056, 1060-1077, 1002 – 1003, 943-947, 530-537, 568-579, 686-691, 1242-1247, 1298 – 1300, 1066-1077, 1052-1056, or 1060-1077 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

A heterologous polypeptide (*e.g.*, adenine deaminase) can be inserted in place of a deleted region of a Cas9 polypeptide. The deleted region can correspond to an N-terminal or C-terminal portion of the Cas9 polypeptide. In some embodiments, the deleted region corresponds to residues 792-872 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the

deleted region corresponds to residues 792-906 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deleted region corresponds to residues 2-791 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deleted region corresponds to residues 1017-1069 as numbered in the above Cas9 reference sequence, or corresponding amino acid residues thereof.

Exemplary internal fusions base editors are provided in Table 3 below:

Table 3: Insertion loci in Cas9 proteins

BE ID	Modification	Other ID
IBE001	Cas9 TadA ins 1015	ISLAY01
IBE002	Cas9 TadA ins 1022	ISLAY02
IBE003	Cas9 TadA ins 1029	ISLAY03
IBE004	Cas9 TadA ins 1040	ISLAY04
IBE005	Cas9 TadA ins 1068	ISLAY05
IBE006	Cas9 TadA ins 1247	ISLAY06
IBE007	Cas9 TadA ins 1054	ISLAY07
IBE008	Cas9 TadA ins 1026	ISLAY08
IBE009	Cas9 TadA ins 768	ISLAY09
IBE020	delta HNH TadA 792	ISLAY20
IBE021	N-term fusion single TadA helix truncated 165-end	ISLAY21
IBE029	TadA-Circular Permutant116 ins1067	ISLAY29
IBE031	TadA- Circular Permutant 136 ins1248	ISLAY31
IBE032	TadA- Circular Permutant 136ins 1052	ISLAY32
IBE035	delta 792-872 TadA ins	ISLAY35
IBE036	delta 792-906 TadA ins	ISLAY36
IBE043	TadA-Circular Permutant 65 ins1246	ISLAY43
IBE044	TadA ins C-term truncate2 791	ISLAY44

A heterologous polypeptide (*e.g.*, deaminase) can be inserted within a structural or functional domain of a Cas9 polypeptide. A heterologous polypeptide (*e.g.*, deaminase) can be inserted between two structural or functional domains of a Cas9 polypeptide. A

heterologous polypeptide (*e.g.*, deaminase) can be inserted in place of a structural or functional domain of a Cas9 polypeptide, for example, after deleting the domain from the Cas9 polypeptide. The structural or functional domains of a Cas9 polypeptide can include, for example, RuvC I, RuvC II, RuvC III, Rec1, Rec2, PI, or HNH.

In some embodiments, the Cas9 polypeptide lacks one or more domains selected from the group consisting of: RuvC I, RuvC II, RuvC III, Rec1, Rec2, PI, or HNH domain. In some embodiments, the Cas9 polypeptide lacks a nuclease domain. In some embodiments, the Cas9 polypeptide lacks an HNH domain. In some embodiments, the Cas9 polypeptide lacks a portion of the HNH domain such that the Cas9 polypeptide has reduced or abolished HNH activity. In some embodiments, the Cas9 polypeptide comprises a deletion of the nuclease domain, and the deaminase is inserted to replace the nuclease domain. In some embodiments, the HNH domain is deleted and the deaminase is inserted in its place. In some embodiments, one or more of the RuvC domains is deleted and the deaminase is inserted in its place.

A fusion protein comprising a heterologous polypeptide can be flanked by a N-terminal and a C-terminal fragment of a napDNAbp. In some embodiments, the fusion protein comprises a deaminase flanked by a N- terminal fragment and a C-terminal fragment of a Cas9 polypeptide. The N terminal fragment or the C terminal fragment can bind the target polynucleotide sequence. The C-terminus of the N terminal fragment or the N-terminus of the C terminal fragment can comprise a part of a flexible loop of a Cas9 polypeptide. The C-terminus of the N terminal fragment or the N-terminus of the C terminal fragment can comprise a part of an alpha-helix structure of the Cas9 polypeptide. The N-terminal fragment or the C-terminal fragment can comprise a DNA binding domain. The N-terminal fragment or the C-terminal fragment can comprise a RuvC domain. The N-terminal fragment or the C-terminal fragment can comprise an HNH domain. In some embodiments, neither of the N-terminal fragment and the C-terminal fragment comprises an HNH domain.

In some embodiments, the C-terminus of the N terminal Cas9 fragment comprises an amino acid that is in proximity to a target nucleobase when the fusion protein deaminates the target nucleobase. In some embodiments, the N-terminus of the C terminal Cas9 fragment comprises an amino acid that is in proximity to a target nucleobase when the fusion protein deaminates the target nucleobase. The insertion location of different deaminases can be different in order to have proximity between the target nucleobase and an amino acid in the C-terminus of the N terminal Cas9 fragment or the N-terminus of the C terminal Cas9 fragment. For example, the insertion position of an deaminase can be at an amino acid

residue selected from the group consisting of: 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, and 1246 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

The N-terminal Cas9 fragment of a fusion protein (*i.e.* the N-terminal Cas9 fragment flanking the deaminase in a fusion protein) can comprise the N-terminus of a Cas9 polypeptide. The N-terminal Cas9 fragment of a fusion protein can comprise a length of at least about: 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, or 1300 amino acids. The N-terminal Cas9 fragment of a fusion protein can comprise a sequence corresponding to amino acid residues: 1-56, 1-95, 1-200, 1-300, 1-400, 1-500, 1-600, 1-700, 1-718, 1-765, 1-780, 1-906, 1-918, or 1-1100 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. The N-terminal Cas9 fragment can comprise a sequence comprising at least: 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity to amino acid residues: 1-56, 1-95, 1-200, 1-300, 1-400, 1-500, 1-600, 1-700, 1-718, 1-765, 1-780, 1-906, 1-918, or 1-1100 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

The C-terminal Cas9 fragment of a fusion protein (*i.e.* the C-terminal Cas9 fragment flanking the deaminase in a fusion protein) can comprise the C-terminus of a Cas9 polypeptide. The C-terminal Cas9 fragment of a fusion protein can comprise a length of at least about: 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, or 1300 amino acids. The C-terminal Cas9 fragment of a fusion protein can comprise a sequence corresponding to amino acid residues: 1099-1368, 918-1368, 906-1368, 780-1368, 765-1368, 718-1368, 94-1368, or 56-1368 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. The N-terminal Cas9 fragment can comprise a sequence comprising at least: 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity to amino acid residues: 1099-1368, 918-1368, 906-1368, 780-1368, 765-1368, 718-1368, 94-1368, or 56-1368 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

The N-terminal Cas9 fragment and C-terminal Cas9 fragment of a fusion protein taken together may not correspond to a full-length naturally occurring Cas9 polypeptide sequence, for example, as set forth in the above Cas9 reference sequence.

The fusion protein described herein can effect targeted deamination with reduced deamination at non-target sites (*e.g.*, off-target sites), such as reduced genome wide spurious deamination. The fusion protein described herein can effect targeted deamination with reduced bystander deamination at non-target sites. The undesired deamination or off-target deamination can be reduced by at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 99% compared with, for example, an end terminus fusion protein comprising the deaminase fused to a N terminus or a C terminus of a Cas9 polypeptide. The undesired deamination or off-target deamination can be reduced by at least one-fold, at least two-fold, at least three-fold, at least four-fold, at least five-fold, at least tenfold, at least fifteen fold, at least twenty fold, at least thirty fold, at least forty fold, at least fifty fold, at least 60 fold, at least 70 fold, at least 80 fold, at least 90 fold, or at least hundred fold, compared with, for example, an end terminus fusion protein comprising the deaminase fused to a N terminus or a C terminus of a Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) of the fusion protein deaminates no more than two nucleobases within the range of an R-loop. In some embodiments, the deaminase of the fusion protein deaminates no more than three nucleobases within the range of the R-loop. In some embodiments, the deaminase of the fusion protein deaminates no more than 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleobases within the range of the R-loop. An R-loop is a three-stranded nucleic acid structure including a DNA:RNA hybrid, a DNA:DNA or an RNA: RNA complementary structure and the associated with single-stranded DNA. As used herein, an R-loop may be formed when a target polynucleotide is contacted with a CRISPR complex or a base editing complex, wherein a portion of a guide polynucleotide, *e.g.* a guide RNA, hybridizes with and displaces with a portion of a target polynucleotide, *e.g.* a target DNA. In some embodiments, an R-loop comprises a hybridized region of a spacer sequence and a target DNA complementary sequence. An R-loop region may be of about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 nucleobase pairs in length. In some embodiments, the R-loop region is about 20 nucleobase pairs in length. It should be understood that, as used herein, an R-loop region is not limited to the target DNA strand that hybridizes with the guide polynucleotide. For example, editing of a target nucleobase within an R-loop region may be to a DNA strand that comprises the complementary strand to a guide RNA, or may be to a DNA strand that is the opposing strand of the strand complementary to the guide RNA. In some embodiments, editing in the region of the R-loop

comprises editing a nucleobase on non-complementary strand (protospacer strand) to a guide RNA in a target DNA sequence.

The fusion protein described herein can effect target deamination in an editing window different from canonical base editing. In some embodiments, a target nucleobase is from about 1 to about 20 bases upstream of a PAM sequence in the target polynucleotide sequence. In some embodiments, a target nucleobase is from about 2 to about 12 bases upstream of a PAM sequence in the target polynucleotide sequence. In some embodiments, a target nucleobase is from about 1 to 9 base pairs, about 2 to 10 base pairs, about 3 to 11 base pairs, about 4 to 12 base pairs, about 5 to 13 base pairs, about 6 to 14 base pairs, about 7 to 15 base pairs, about 8 to 16 base pairs, about 9 to 17 base pairs, about 10 to 18 base pairs, about 11 to 19 base pairs, about 12 to 20 base pairs, about 1 to 7 base pairs, about 2 to 8 base pairs, about 3 to 9 base pairs, about 4 to 10 base pairs, about 5 to 11 base pairs, about 6 to 12 base pairs, about 7 to 13 base pairs, about 8 to 14 base pairs, about 9 to 15 base pairs, about 10 to 16 base pairs, about 11 to 17 base pairs, about 12 to 18 base pairs, about 13 to 19 base pairs, about 14 to 20 base pairs, about 1 to 5 base pairs, about 2 to 6 base pairs, about 3 to 7 base pairs, about 4 to 8 base pairs, about 5 to 9 base pairs, about 6 to 10 base pairs, about 7 to 11 base pairs, about 8 to 12 base pairs, about 9 to 13 base pairs, about 10 to 14 base pairs, about 11 to 15 base pairs, about 12 to 16 base pairs, about 13 to 17 base pairs, about 14 to 18 base pairs, about 15 to 19 base pairs, about 16 to 20 base pairs, about 1 to 3 base pairs, about 2 to 4 base pairs, about 3 to 5 base pairs, about 4 to 6 base pairs, about 5 to 7 base pairs, about 6 to 8 base pairs, about 7 to 9 base pairs, about 8 to 10 base pairs, about 9 to 11 base pairs, about 10 to 12 base pairs, about 11 to 13 base pairs, about 12 to 14 base pairs, about 13 to 15 base pairs, about 14 to 16 base pairs, about 15 to 17 base pairs, about 16 to 18 base pairs, about 17 to 19 base pairs, about 18 to 20 base pairs away or upstream of the PAM sequence. In some embodiments, a target nucleobase is about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more base pairs away from or upstream of the PAM sequence. In some embodiments, a target nucleobase is about 1, 2, 3, 4, 5, 6, 7, 8, or 9 base pairs upstream of the PAM sequence. In some embodiments, a target nucleobase is about 2, 3, 4, or 6 base pairs upstream of the PAM sequence.

The fusion protein can comprise more than one heterologous polypeptide. For example, the fusion protein can additionally comprise one or more UGI domains and/or one or more nuclear localization signals. The two or more heterologous domains can be inserted in tandem. The two or more heterologous domains can be inserted at locations such that they are not in tandem in the NapDNAbp.

A fusion protein can comprise a linker between the deaminase and the napDNAbp polypeptide. The linker can be a peptide or a non-peptide linker. For example, the linker can be an XTEN, (GGGS)n (SEQ ID NO: 334), (GGGGS)n (SEQ ID NO: 335), (G)n, (EAAAK)n (SEQ ID NO: 336), (GGS)n, SGSETPGTSESATPES (SEQ ID NO: 337). In some embodiments, the fusion protein comprises a linker between the N-terminal Cas9 fragment and the deaminase. In some embodiments, the fusion protein comprises a linker between the C-terminal Cas9 fragment and the deaminase. In some embodiments, the N-terminal and C-terminal fragments of napDNAbp are connected to the deaminase with a linker. In some embodiments, the N-terminal and C-terminal fragments are joined to the deaminase domain without a linker. In some embodiments, the fusion protein comprises a linker between the N-terminal Cas9 fragment and the deaminase, but does not comprise a linker between the C-terminal Cas9 fragment and the deaminase. In some embodiments, the fusion protein comprises a linker between the C-terminal Cas9 fragment and the deaminase, but does not comprise a linker between the N-terminal Cas9 fragment and the deaminase.

In some embodiments, the napDNAbp in the fusion protein is a Cas12 polypeptide, *e.g.*, Cas12b/C2c1, or a fragment thereof. The Cas12 polypeptide can be a variant Cas12 polypeptide. In other embodiments, the N- or C-terminal fragments of the Cas12 polypeptide comprise a nucleic acid programmable DNA binding domain or a RuvC domain. In other embodiments, the fusion protein contains a linker between the Cas12 polypeptide and the catalytic domain. In other embodiments, the amino acid sequence of the linker is GGSGGS (SEQ ID NO: 338) or GSSGSETPGTSESATPESSG (SEQ ID NO: 339). In other embodiments, the linker is a rigid linker. In other embodiments of the above aspects, the linker is encoded by GGAGGCTCTGGAGGAAGC (SEQ ID NO: 340) or GGCTCTTCTGGATCTGAAACACCTGGCACAAGCGAGAGCGCCACCCCTGAGAGC TCTGGC (SEQ ID NO: 341).

Fusion proteins comprising a heterologous catalytic domain flanked by N- and C-terminal fragments of a Cas12 polypeptide are also useful for base editing in the methods as described herein. Fusion proteins comprising Cas12 and one or more deaminase domains, *e.g.*, adenosine deaminase, or comprising an adenosine deaminase domain flanked by Cas12 sequences are also useful for highly specific and efficient base editing of target sequences. In an embodiment, a chimeric Cas12 fusion protein contains a heterologous catalytic domain (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) inserted within a Cas12 polypeptide. In some embodiments, the fusion protein comprises an adenosine deaminase domain and a cytidine deaminase domain inserted within

a Cas12. In some embodiments, an adenosine deaminase is fused within Cas12 and a cytidine deaminase is fused to the C-terminus. In some embodiments, an adenosine deaminase is fused within Cas12 and a cytidine deaminase fused to the N-terminus. In some embodiments, a cytidine deaminase is fused within Cas12 and an adenosine deaminase is fused to the C-terminus. In some embodiments, a cytidine deaminase is fused within Cas12 and an adenosine deaminase fused to the N-terminus. Exemplary structures of a fusion protein with an adenosine deaminase and a cytidine deaminase and a Cas12 are provided as follows:

NH2-[Cas12(adenosine deaminase)]-[cytidine deaminase]-COOH;

NH2-[cytidine deaminase]-[Cas12(adenosine deaminase)]-COOH;

NH2-[Cas12(cytidine deaminase)]-[adenosine deaminase]-COOH; or

NH2-[adenosine deaminase]-[Cas12(cytidine deaminase)]-COOH;

In some embodiments, the "-" used in the general architecture above indicates the presence of an optional linker.

In various embodiments, the catalytic domain has DNA modifying activity (*e.g.*, deaminase activity), such as adenosine deaminase activity. In some embodiments, the adenosine deaminase is a TadA (*e.g.*, TadA*7.10). In some embodiments, the TadA is a TadA*8. In some embodiments, a TadA*8 is fused within Cas12 and a cytidine deaminase is fused to the C-terminus. In some embodiments, a TadA*8 is fused within Cas12 and a cytidine deaminase fused to the N-terminus. In some embodiments, a cytidine deaminase is fused within Cas12 and a TadA*8 is fused to the C-terminus. In some embodiments, a cytidine deaminase is fused within Cas12 and a TadA*8 fused to the N-terminus. Exemplary structures of a fusion protein with a TadA*8 and a cytidine deaminase and a Cas12 are provided as follows:

N-[Cas12(TadA*8)]-[cytidine deaminase]-C;

N-[cytidine deaminase]-[Cas12(TadA*8)]-C;

N-[Cas12(cytidine deaminase)]-[TadA*8]-C; or

N-[TadA*8]-[Cas12(cytidine deaminase)]-C.

In some embodiments, the "-" used in the general architecture above indicates the presence of an optional linker.

In other embodiments, the fusion protein contains one or more catalytic domains. In other embodiments, at least one of the one or more catalytic domains is inserted within the Cas12 polypeptide or is fused at the Cas12 N- terminus or C-terminus. In other embodiments, at least one of the one or more catalytic domains is inserted within a loop, an

alpha helix region, an unstructured portion, or a solvent accessible portion of the Cas12 polypeptide. In other embodiments, the Cas12 polypeptide is Cas12a, Cas12b, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ. In other embodiments, the Cas12 polypeptide has at least about 85% amino acid sequence identity to *Bacillus hisashii* Cas12b, *Bacillus thermoamylovorans* Cas12b, *Bacillus* sp. V3-13 Cas12b, or *Alicyclobacillus acidiphilus* Cas12b (SEQ ID NO: 342). In other embodiments, the Cas12 polypeptide has at least about 90% amino acid sequence identity to *Bacillus hisashii* Cas12b (SEQ ID NO: 343), *Bacillus thermoamylovorans* Cas12b, *Bacillus* sp. V3-13 Cas12b, or *Alicyclobacillus acidiphilus* Cas12b. In other embodiments, the Cas12 polypeptide has at least about 95% amino acid sequence identity to *Bacillus hisashii* Cas12b, *Bacillus thermoamylovorans* Cas12b (SEQ ID NO: 344), *Bacillus* sp. V3-13 Cas12b (SEQ ID NO: 345), or *Alicyclobacillus acidiphilus* Cas12b. In other embodiments, the Cas12 polypeptide contains or consists essentially of a fragment of *Bacillus hisashii* Cas12b, *Bacillus thermoamylovorans* Cas12b, *Bacillus* sp. V3-13 Cas12b, or *Alicyclobacillus acidiphilus* Cas12b. In embodiments, the Cas12 polypeptide contains BvCas12b (V4), which in some embodiments is expressed as 5' mRNA Cap---5' UTR---bhCas12b---STOP sequence --- 3' UTR --- 120polyA tail (SEQ ID NOs: 346-348).

In other embodiments, the catalytic domain is inserted between amino acid positions 153-154, 255-256, 306-307, 980-981, 1019-1020, 534-535, 604-605, or 344-345 of BhCas12b or a corresponding amino acid residue of Cas12a, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ. In other embodiments, the catalytic domain is inserted between amino acids P153 and S154 of BhCas12b. In other embodiments, the catalytic domain is inserted between amino acids K255 and E256 of BhCas12b. In other embodiments, the catalytic domain is inserted between amino acids D980 and G981 of BhCas12b. In other embodiments, the catalytic domain is inserted between amino acids K1019 and L1020 of BhCas12b. In other embodiments, the catalytic domain is inserted between amino acids F534 and P535 of BhCas12b. In other embodiments, the catalytic domain is inserted between amino acids K604 and G605 of BhCas12b. In other embodiments, the catalytic domain is inserted between amino acids H344 and F345 of BhCas12b. In other embodiments, catalytic domain is inserted between amino acid positions 147 and 148, 248 and 249, 299 and 300, 991 and 992, or 1031 and 1032 of BvCas12b or a corresponding amino acid residue of Cas12a, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ. In other embodiments, the catalytic domain is inserted between

amino acids P147 and D148 of BvCas12b. In other embodiments, the catalytic domain is inserted between amino acids G248 and G249 of BvCas12b. In other embodiments, the catalytic domain is inserted between amino acids P299 and E300 of BvCas12b. In other embodiments, the catalytic domain is inserted between amino acids G991 and E992 of BvCas12b. In other embodiments, the catalytic domain is inserted between amino acids K1031 and M1032 of BvCas12b. In other embodiments, the catalytic domain is inserted between amino acid positions 157 and 158, 258 and 259, 310 and 311, 1008 and 1009, or 1044 and 1045 of AaCas12b or a corresponding amino acid residue of Cas12a, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ. In other embodiments, the catalytic domain is inserted between amino acids P157 and G158 of AaCas12b. In other embodiments, the catalytic domain is inserted between amino acids V258 and G259 of AaCas12b. In other embodiments, the catalytic domain is inserted between amino acids D310 and P311 of AaCas12b. In other embodiments, the catalytic domain is inserted between amino acids G1008 and E1009 of AaCas12b. In other embodiments, the catalytic domain is inserted between amino acids G1044 and K1045 at of AaCas12b.

In other embodiments, the fusion protein contains a nuclear localization signal (*e.g.*, a bipartite nuclear localization signal). In other embodiments, the amino acid sequence of the nuclear localization signal is MAPKKKRKVGIHGVPAA (SEQ ID NO: 349). In other embodiments of the above aspects, the nuclear localization signal is encoded by the following sequence:
ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCC (SEQ ID NO: 350). In other embodiments, the Cas12b polypeptide contains a mutation that silences the catalytic activity of a RuvC domain. In other embodiments, the Cas12b polypeptide contains D574A, D829A and/or D952A mutations. In other embodiments, the fusion protein further contains a tag (*e.g.*, an influenza hemagglutinin tag).

In some embodiments, the fusion protein comprises a napDNAbp domain (*e.g.*, Cas12-derived domain) with an internally fused nucleobase editing domain (*e.g.*, all or a portion of a deaminase domain, *e.g.*, an adenosine deaminase domain). In some embodiments, the napDNAbp is a Cas12b. In some embodiments, the base editor comprises a BhCas12b domain with an internally fused TadA*8 domain inserted at the loci provided in Table 4 below.

Table 4: Insertion loci in Cas12b proteins

BhCas12b	Insertion site	Inserted between aa
position 1	153	PS
position 2	255	KE
position 3	306	DE
position 4	980	DG
position 5	1019	KL
position 6	534	FP
position 7	604	KG
position 8	344	HF
BvCas12b	Insertion site	Inserted between aa
position 1	147	PD
position 2	248	GG
position 3	299	PE
position 4	991	GE
position 5	1031	KM
AaCas12b	Insertion site	Inserted between aa
position 1	157	PG
position 2	258	VG
position 3	310	DP
position 4	1008	GE
position 5	1044	GK

By way of nonlimiting example, an adenosine deaminase (*e.g.*, TadA*8.13) may be inserted into a BhCas12b to produce a fusion protein (*e.g.*, TadA*8.13-BhCas12b) that effectively edits a nucleic acid sequence.

In some embodiments, the base editing system described herein is an ABE with TadA inserted into a Cas9. Polypeptide sequences of relevant ABEs with TadA inserted into a Cas9 are provided in the attached Sequence Listing as SEQ ID NOs: 351-396.

In some embodiments, adenosine base editors were generated to insert TadA or variants thereof into the Cas9 polypeptide at the identified positions.

Exemplary, yet nonlimiting, fusion proteins are described in International PCT Application Nos. PCT/US2020/016285 and U.S. Provisional Application Nos. 62/852,228 and 62/852,224, the contents of which are incorporated by reference herein in their entireties.

A to G Editing

In some embodiments, a base editor described herein comprises an adenosine deaminase domain. Such an adenosine deaminase domain of a base editor can facilitate the editing of an adenine (A) nucleobase to a guanine (G) nucleobase by deaminating the A to form inosine (I), which exhibits base pairing properties of G. Adenosine deaminase is capable of deaminating (*i.e.*, removing an amine group) adenine of a deoxyadenosine residue in deoxyribonucleic acid (DNA). In some embodiments, an A-to-G base editor further comprises an inhibitor of inosine base excision repair, for example, a uracil glycosylase inhibitor (UGI) domain or a catalytically inactive inosine specific nuclease. Without wishing to be bound by any particular theory, the UGI domain or catalytically inactive inosine specific nuclease can inhibit or prevent base excision repair of a deaminated adenosine residue (*e.g.*, inosine), which can improve the activity or efficiency of the base editor.

A base editor comprising an adenosine deaminase can act on any polynucleotide, including DNA, RNA and DNA-RNA hybrids. In certain embodiments, a base editor comprising an adenosine deaminase can deaminate a target A of a polynucleotide comprising RNA. For example, the base editor can comprise an adenosine deaminase domain capable of deaminating a target A of an RNA polynucleotide and/or a DNA-RNA hybrid polynucleotide. In an embodiment, an adenosine deaminase incorporated into a base editor comprises all or a portion of adenosine deaminase acting on RNA (ADAR, *e.g.*, ADAR1 or ADAR2) or tRNA (ADAT). A base editor comprising an adenosine deaminase domain can also be capable of deaminating an A nucleobase of a DNA polynucleotide. In an embodiment an adenosine deaminase domain of a base editor comprises all or a portion of an ADAT comprising one or more mutations which permit the ADAT to deaminate a target A in DNA. For example, the base editor can comprise all or a portion of an ADAT from Escherichia coli (EcTadA) comprising one or more of the following mutations: D108N, A106V, D147Y, E155V, L84F, H123Y, I156F, or a corresponding mutation in another adenosine deaminase. Exemplary ADAT homolog polypeptide sequences are provided in the Sequence Listing as SEQ ID NOs: 1, 397-403.

The adenosine deaminase can be derived from any suitable organism (*e.g.*, *E. coli*). In some embodiments, the adenosine deaminase is from a prokaryote. In some embodiments, the adenosine deaminase is from a bacterium. In some embodiments, the adenosine deaminase is from *Escherichia coli*, *Staphylococcus aureus, Salmonella typhi, Shewanella putrefaciens, Haemophilus influenzae, Caulobacter crescentus, or Bacillus subtilis*. In some embodiments, the adenosine deaminase is from *E. coli*. In some embodiments, the adenine deaminase is a naturally-occurring adenosine deaminase that includes one or more mutations corresponding to any of the mutations provided herein (*e.g.*, mutations in ecTadA). The corresponding residue in any homologous protein can be identified by *e.g.*, sequence alignment and determination of homologous residues. The mutations in any naturally-occurring adenosine deaminase (*e.g.*, having homology to ecTadA) that correspond to any of the mutations described herein (*e.g.*, any of the mutations identified in ecTadA) can be generated accordingly.

In some embodiments, the adenosine deaminase comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of the amino acid sequences set forth in any of the adenosine deaminases provided herein. It should be appreciated that adenosine deaminases provided herein may include one or more mutations (*e.g.*, any of the mutations provided herein). The disclosure provides any deaminase domains with a certain percent identify plus any of the mutations or combinations thereof described herein. In some embodiments, the adenosine deaminase comprises an amino acid sequence that has 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, or more mutations compared to a reference sequence, or any of the adenosine deaminases provided herein. In some embodiments, the adenosine deaminase comprises an amino acid sequence that has at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 110, at least 120, at least 130, at least 140, at least 150, at least 160, or at least 170 identical contiguous amino acid residues as compared to any one of the amino acid sequences known in the art or described herein.

It should be appreciated that any of the mutations provided herein (*e.g.*, based on the TadA reference sequence) can be introduced into other adenosine deaminases, such as *E. coli* TadA (ecTadA), *S. aureus* TadA (saTadA), or other adenosine deaminases (*e.g.*, bacterial adenosine deaminases). It would be apparent to the skilled artisan that additional deaminases

may similarly be aligned to identify homologous amino acid residues that can be mutated as provided herein. Thus, any of the mutations identified in the TadA reference sequence can be made in other adenosine deaminases (*e.g.*, ecTada) that have homologous amino acid residues. It should also be appreciated that any of the mutations provided herein can be made individually or in any combination in the TadA reference sequence or another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises a D108X mutation in the TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a D108G, D108N, D108V, D108A, or D108Y mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase. It should be appreciated, however, that additional deaminases may similarly be aligned to identify homologous amino acid residues that can be mutated as provided herein.

In some embodiments, the adenosine deaminase comprises an A106X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an A106V mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises a E155X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a E155D, E155G, or E155V mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises a D147X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a D147Y, mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an A106X, E155X, or D147X, mutation in the TadA reference sequence, or a corresponding mutation in another

adenosine deaminase (*e.g.*, ecTadA), where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an E155D, E155G, or E155V mutation. In some embodiments, the adenosine deaminase comprises a D147Y.

It should also be appreciated that any of the mutations provided herein may be made individually or in any combination in ecTadA or another adenosine deaminase. For example, an adenosine deaminase may contain a D108N, a A106V, a E155V, and/or a D147Y mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, an adenosine deaminase comprises the following group of mutations (groups of mutations are separated by a ";") in TadA reference sequence, or corresponding mutations in another adenosine deaminase: D108N and A106V; D108N and E155V; D108N and D147Y; A106V and E155V; A106V and D147Y; E155V and D147Y; D108N, A106V, and E155V; D108N, A106V, and D147Y; D108N, E155V, and D147Y; A106V, E155V, and D147Y; and D108N, A106V, E155V, and D147Y. It should be appreciated, however, that any combination of corresponding mutations provided herein may be made in an adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises a combination of mutations in a TadA reference sequence (e.g., TadA*7.10), or corresponding mutations in another adenosine deaminase: V82G + Y147T + Q154S; I76Y + V82G + Y147T + Q154S; L36H + V82G + Y147T + Q154S + N157K; V82G + Y147D + F149Y + Q154S + D167N; L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N; L36H + I76Y + V82G + Y147T + Q154S + N157K; I76Y + V82G + Y147D + F149Y + Q154S + D167N; or L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N.

In some embodiments, the adenosine deaminase comprises one or more of a H8X, T17X, L18X, W23X, L34X, W45X, R51X, A56X, E59X, E85X, M94X, I95X, V102X, F104X, A106X, R107X, D108X, K110X, M118X, N127X, A138X, F149X, M151X, R153X, Q154X, I156X, and/or K157X mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of H8Y, T17S, L18E, W23L, L34S, W45L, R51H, A56E, or A56S, E59G, E85K, or E85G, M94L, I95L, V102A, F104L, A106V, R107C, or R107H, or R107P, D108G, or D108N, or D108V, or D108A, or D108Y, K110I, M118K, N127S, A138V, F149Y, M151V, R153C, Q154L,

I156D, and/or K157R mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one or more of a H8X, D108X, and/or N127X mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where X indicates the presence of any amino acid. In some embodiments, the adenosine deaminase comprises one or more of a H8Y, D108N, and/or N127S mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one or more of H8X, R26X, M61X, L68X, M70X, A106X, D108X, A109X, N127X, D147X, R152X, Q154X, E155X, K161X, Q163X, and/or T166X mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of H8Y, R26W, M61I, L68Q, M70V, A106T, D108N, A109T, N127S, D147Y, R152C, Q154H or Q154R, E155G or E155V or E155D, K161Q, Q163H, and/or T166P mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of H8X, D108X, N127X, D147X, R152X, and Q154X in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA), where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, seven, or eight mutations selected from the group consisting of H8X, M61X, M70X, D108X, N127X, Q154X, E155X, and Q163X in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA), where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, or five, mutations selected from the group consisting of H8X, D108X, N127X, E155X, and T166X in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA), where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of H8X, A106X, and D108X, or a

corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, seven, or eight mutations selected from the group consisting of H8X, R26X, L68X, D108X, N127X, D147X, and E155X, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, or seven mutations selected from the group consisting of H8X, R126X, L68X, D108X, N127X, D147X, and E155X in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, or five mutations selected from the group consisting of H8X, D108X, A109X, N127X, and E155X in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of H8Y, D108N, N127S, D147Y, R152C, and Q154H in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, seven, or eight mutations selected from the group consisting of H8Y, M61I, M70V, D108N, N127S, Q154R, E155G and Q163H in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises one, two, three, four, or five, mutations selected from the group consisting of H8Y, D108N, N127S, E155V, and T166P in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of H8Y, A106T, D108N, N127S, E155D, and K161Q in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, seven, or eight mutations selected from the group consisting of H8Y, R26W, L68Q, D108N, N127S, D147Y, and E155V in TadA reference sequence, or a corresponding

mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises one, two, three, four, or five, mutations selected from the group consisting of H8Y, D108N, A109T, N127S, and E155G in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises one or more of the or one or more corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a D108N, D108G, or D108V mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a A106V and D108N mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises R107C and D108N mutations in TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a H8Y, D108N, N127S, D147Y, and Q154H mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a H8Y, D108N, N127S, D147Y, and E155V mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a D108N, D147Y, and E155V mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a H8Y, D108N, and N127S mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a A106V, D108N, D147Y, and E155V mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises one or more of S2X, H8X, I49X, L84X, H123X, N127X, I156X, and/or K160X mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of S2A, H8Y, I49F, L84F, H123Y, N127S, I156F, and/or K160S mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an L84X mutation adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an L84F mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an H123X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an H123Y mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an I156X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an I156F mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, or seven mutations selected from the group consisting of L84X, A106X, D108X, H123X, D147X, E155X, and I156X in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of S2X, I49X, A106X, D108X, D147X, and E155X in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, or five mutations selected from the group consisting of H8X, A106X, D108X, N127X, and K160X in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, or seven mutations selected from the group consisting of L84F, A106V, D108N, H123Y,

D147Y, E155V, and I156F in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of S2A, I49F, A106V, D108N, D147Y, and E155V in TadA reference sequence.

In some embodiments, the adenosine deaminase comprises one, two, three, four, or five mutations selected from the group consisting of H8Y, A106T, D108N, N127S, and K160S in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one or more of a E25X, R26X, R107X, A142X, and/or A143X mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of E25M, E25D, E25A, E25R, E25V, E25S, E25Y, R26G, R26N, R26Q, R26C, R26L, R26K, R107P, R107K, R107A, R107N, R107W, R107H, R107S, A142N, A142D, A142G, A143D, A143G, A143E, A143L, A143W, A143M, A143S, A143Q, and/or A143R mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of the mutations described herein corresponding to TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an E25X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an E25M, E25D, E25A, E25R, E25V, E25S, or E25Y mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an R26X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises R26G, R26N, R26Q, R26C, R26L, or R26K mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an R107X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase,

where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an R107P, R107K, R107A, R107N, R107W, R107H, or R107S mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an A142X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an A142N, A142D, A142G, mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an A143X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an A143D, A143G, A143E, A143L, A143W, A143M, A143S, A143Q, and/or A143R mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises one or more of a H36X, N37X, P48X, I49X, R51X, M70X, N72X, D77X, E134X, S146X, Q154X, K157X, and/or K161X mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of H36L, N37T, N37S, P48T, P48L, I49V, R51H, R51L, M70L, N72S, D77G, E134G, S146R, S146C, Q154H, K157N, and/or K161T mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an H36X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an H36L mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an N37X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase,

where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an N37T or N37S mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an P48X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an P48T or P48L mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an R51X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an R51H or R51L mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an S146X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an S146R or S146C mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an K157X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a K157N mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an P48X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a P48S, P48T, or P48A mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an A142X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a A142N mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an W23X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a W23R or W23L mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an R152X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a R152P or R52H mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In one embodiment, the adenosine deaminase may comprise the mutations H36L, R51L, L84F, A106V, D108N, H123Y, S146C, D147Y, E155V, I156F, and K157N. In some embodiments, the adenosine deaminase comprises the following combination of mutations relative to TadA reference sequence, where each mutation of a combination is separated by a "_" and each combination of mutations is between parentheses:

(A106V_D108N),
(R107C_D108N),
(H8Y_D108N_N127S_D147Y_Q154H),
(H8Y _D108N_N127S_D147Y_E155V),
(D108N_D147Y_E155V),
(H8Y_D108N_N127S),
(H8Y_D108N_N127S_D147Y_Q154H),
(A106V_D108N_D147Y_E155V),
(D108Q_D147Y_E155V),
(D108M_D147Y_E155V),
(D108L_D147Y_E155V),
(D108K_D147Y_E155V),

(D108I_D147Y_E155V),
(D108F_D147Y_E155V),
(A106V_D108N_D147Y),
(A106V_D108M_D147Y_E155V),
(E59A_A106V_D108N_D147Y_E155V),
(E59A cat dead_A106V_D108N_D147Y_E155V),
(L84F_A106V_D108N_H123Y_D147Y_E155V_I156Y),
(L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(D103A_D104N),
(G22P_D103A_D104N),
(D103A_D104N_S138A),
(R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),
(E25G_R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),
(E25D_R26G_L84F_A106V_R107K_D108N_H123Y_A142N_A143G_D147Y_E155V_I156F),
(R26Q_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),
(E25M_R26G_L84F_A106V_R107P_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),
(R26C_L84F_A106V_R107H_D108N_H123Y_A142N_D147Y_E155V_I156F),
(L84F_A106V_D108N_H123Y_A142N_A143L_D147Y_E155V_I156F),
(R26G_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),
(E25A_R26G_L84F_A106V_R107N_D108N_H123Y_A142N_A143E_D147Y_E155V_I156F),
(R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),
(A106V_D108N_A142N_D147Y_E155V),
(R26G_A106V_D108N_A142N_D147Y_E155V),
(E25D_R26G_A106V_R107K_D108N_A142N_A143G_D147Y_E155V),
(R26G_A106V_D108N_R107H_A142N_A143D_D147Y_E155V),
(E25D_R26G_A106V_D108N_A142N_D147Y_E155V),
(A106V_R107K_D108N_A142N_D147Y_E155V),
(A106V_D108N_A142N_A143G_D147Y_E155V),
(A106V_D108N_A142N_A143L_D147Y_E155V),
(H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N),
(N37T_P48T_M70L_L84F_A106V_D108N_H123Y_D147Y_I49V_E155V_I156F),
(N37S_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K161T),
(H36L_L84F_A106V_D108N_H123Y_D147Y_Q154H_E155V_I156F),
(N72S_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F),
(H36L_P48L_L84F_A106V_D108N_H123Y_E134G_D147Y_E155V_I156F),
(H36L_L84F_A106V_D108N_H123Y_ D147Y_E155V_I156F_K157N)
(H36L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F),

(L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T),
(N37S_R51H_D77G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(R51L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N),
(D24G_Q71R_L84F_H96L_A106V_D108N_H123Y_D147Y_E155V_I156F_K160E),
(H36L_G67V_L84F_A106V_D108N_H123Y_S146T_D147Y_E155V_I156F),
(Q71L_L84F_A106V_D108N_H123Y_L137M_A143E_D147Y_E155V_I156F),
(E25G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L),
(L84F_A91T_F104I_A106V_D108N_H123Y_D147Y_E155V_I156F),
(N72D_L84F_A106V_D108N_H123Y_G125A_D147Y_E155V_I156F),
(P48S_L84F_S97C_A106V_D108N_H123Y_D147Y_E155V_I156F),
(W23G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(D24G_P48L_Q71R_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L),
(L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),
(H36L_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N),
(N37S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_K161T),
(L84F_A106V_D108N_D147Y_E155V_I156F),
(R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K161T),
(L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K161T),
(L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E_K161T),
(L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E),
(R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(R74A_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(L84F_R98Q_A106V_D108N_H123Y_D147Y_E155V_I156F),
(L84F_A106V_D108N_H123Y_R129Q_D147Y_E155V_I156F),
(P48S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),
(P48S_A142N),
(P48T_I49V_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_L157N),
(P48T_I49V_A142N),
(H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N),
(H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F
(H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N),
(H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_ I156F
_K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N),

(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F _K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F _K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N),
(W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F _K161T),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152H_E155V_I156F _K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F _K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V _I156F _K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_E155V_I156F _K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_R152P _E155V_I156F _K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F _K161T),
(W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V _I156F _K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_R152P_E155 V_I156F _K157N).

In some embodiments, the TadA deaminase is TadA variant. In some embodiments, the TadA variant is TadA*7.10. In particular embodiments, the fusion proteins comprise a single TadA*7.10 domain (*e.g.*, provided as a monomer). In other embodiments, the fusion protein comprises TadA*7.10 and TadA(wt), which are capable of forming heterodimers. In one embodiment, a fusion protein of the invention comprises a wild-type TadA linked to TadA*7.10, which is linked to Cas9 nickase.

In some embodiments, TadA*7.10 comprises at least one alteration. In some embodiments, the adenosine deaminase comprises an alteration in the following sequence: TadA*7.10

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 1)

In some embodiments, TadA*7.10 comprises an alteration at amino acid 82 and/or 166. In particular embodiments, TadA*7.10 comprises one or more of the following alterations: Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R. In other embodiments, a variant of TadA*7.10 comprises a combination of alterations selected from the group of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R +Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R.

In some embodiments, a variant of TadA*7.10 comprises one or more of alterations selected from the group of L36H, I76Y, V82G, Y147T, Y147D, F149Y, Q154S, N157K, and/or D167N. In some embodiments, a variant of TadA*7.10 comprises V82G, Y147T/D, Q154S, and one or more of L36H, I76Y, F149Y, N157K, and D167N. In other embodiments, a variant of TadA*7.10 comprises a combination of alterations selected from the group of: V82G + Y147T + Q154S; I76Y + V82G + Y147T + Q154S; L36H + V82G + Y147T + Q154S + N157K; V82G + Y147D + F149Y + Q154S + D167N; L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N; L36H + I76Y + V82G + Y147T + Q154S + N157K; I76Y + V82G + Y147D + F149Y + Q154S + D167N; L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N.

In some embodiments, an adenosine deaminase variant (*e.g.*, TadA*8) comprises a deletion. In some embodiments, an adenosine deaminase variant comprises a deletion of the C terminus. In particular embodiments, an adenosine deaminase variant comprises a deletion of the C terminus beginning at residue 149, 150, 151, 152, 153, 154, 155, 156, and 157, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, an adenosine deaminase variant (*e.g.*, TadA*8) is a monomer comprising one or more of the following alterations: Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant (TadA*8) is a monomer comprising a combination of alterations selected from the group of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S +

Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R +Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, the adenosine deaminase variant is a homodimer comprising two adenosine deaminase domains (*e.g.*, TadA*8) each having one or more of the following alterations Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a homodimer comprising two adenosine deaminase domains (*e.g.*, TadA*8) each having a combination of alterations selected from the group of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R +Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, a base editor of the disclosure comprising an adenosine deaminase variant (*e.g.*, TadA*8) monomer comprising one or more of the following alterations: R26C, V88A, A109S, T111R, D119N, H122N, Y147D, F149Y, T166I and/or D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant (TadA*8) monomer comprises a combination of alterations selected from the group of: R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N; V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; V88A + T111R + D119N + F149Y; and A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In some embodiments, an adenosine deaminase variant (*e.g.*, MSP828) is a monomer comprising one or more of the following alterations L36H, I76Y, V82G, Y147T, Y147D, F149Y, Q154S, N157K, and/or D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In some embodiments, an adenosine deaminase variant (*e.g.*, MSP828) is a monomer comprising V82G, Y147T/D, Q154S, and one or more of L36H, I76Y, F149Y, N157K, and D167N, relative to TadA*7.10, the TadA

reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant (TadA variant) is a monomer comprising a combination of alterations selected from the group of: V82G + Y147T + Q154S; I76Y + V82G + Y147T + Q154S; L36H + V82G + Y147T + Q154S + N157K; V82G + Y147D + F149Y + Q154S + D167N; L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N; L36H + I76Y + V82G + Y147T + Q154S + N157K; I76Y + V82G + Y147D + F149Y + Q154S + D167N; L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, the adenosine deaminase variant is a heterodimer of a wild-type adenosine deaminase domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising one or more of the following alterations Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a heterodimer of a wild-type adenosine deaminase domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising a combination of alterations selected from the group of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R +Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, a base editor of the disclosure comprising an adenosine deaminase variant (*e.g.*, TadA*8) homodimer comprising two adenosine deaminase domains (*e.g.*, TadA*8) each having one or more of the following alterations R26C, V88A, A109S, T111R, D119N, H122N, Y147D, F149Y, T166I and/or D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a homodimer comprising two adenosine deaminase domains (*e.g.*, TadA*8) each having a combination of alterations selected from the group of: R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N; V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; V88A + T111R + D119N + F149Y; and A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In some embodiments, an adenosine deaminase variant is a homodimer comprising two adenosine deaminase domains (*e.g.*, TadA*7.10) each having one or more of the following alterations L36H, I76Y, V82G, Y147T, Y147D, F149Y, Q154S, N157K, and/or D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In some embodiments, an adenosine deaminase variant is a homodimer comprising two adenosine deaminase variant domains (*e.g.*, MSP828) each having the following alterations V82G, Y147T/D, Q154S, and one or more of L36H, I76Y, F149Y, N157K, and D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a homodimer comprising two adenosine deaminase domains (*e.g.*, TadA*7.10) each having a combination of alterations selected from the group of: V82G + Y147T + Q154S; I76Y + V82G + Y147T + Q154S; L36H + V82G + Y147T + Q154S + N157K; V82G + Y147D + F149Y + Q154S + D167N; L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N; L36H + I76Y + V82G + Y147T + Q154S + N157K; I76Y + V82G + Y147D + F149Y + Q154S + D167N; L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, the adenosine deaminase variant is a heterodimer of a TadA*7.10 domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising one or more of the following alterations Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a heterodimer of a TadA*7.10 domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising a combination of alterations selected from the group of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R +Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, a base editor comprises a heterodimer of a wild-type adenosine deaminase domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising one or more of the following alterations R26C, V88A, A109S, T111R, D119N, H122N, Y147D, F149Y, T166I and/or D167N, relative to TadA*7.10, the TadA reference sequence,

or a corresponding mutation in another TadA. In other embodiments, the base editor comprises a heterodimer of a wild-type adenosine deaminase domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising a combination of alterations selected from the group of: R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N; V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; V88A + T111R + D119N + F149Y; and A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, the adenosine deaminase variant is a heterodimer of a wild-type adenosine deaminase domain and an adenosine deaminase variant domain (*e.g.*, TadA*7.10) comprising one or more of the following alterations L36H, I76Y, V82G, Y147T, Y147D, F149Y, Q154S, N157K, and/or D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In some embodiments, an adenosine deaminase variant is a heterodimer comprising a wild-type adenosine deaminase domain and an adenosine deaminase variant domain (*e.g.*, MSP828) having the following alterations V82G, Y147T/D, Q154S, and one or more of L36H, I76Y, F149Y, N157K, and D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a heterodimer of a wild-type adenosine deaminase domain and an adenosine deaminase variant domain (*e.g.*, TadA*7.10) comprising a combination of alterations selected from the group of: V82G + Y147T + Q154S; I76Y + V82G + Y147T + Q154S; L36H + V82G + Y147T + Q154S + N157K; V82G + Y147D + F149Y + Q154S + D167N; L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N; L36H + I76Y + V82G + Y147T + Q154S + N157K; I76Y + V82G + Y147D + F149Y + Q154S + D167N; L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, the adenosine deaminase variant is a heterodimer of a TadA*7.10 domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising one or more of the following alterations Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a heterodimer of a TadA*7.10 domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising a combination of alterations selected from the group of: Y147T +

Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R +Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In particular embodiments, an adenosine deaminase heterodimer comprises a TadA*8 domain and an adenosine deaminase domain selected from *Staphylococcus aureus* (*S. aureus*) TadA, *Bacillus subtilis* (*B. subtilis*) TadA, *Salmonella typhimurium* (*S. typhimurium*) TadA, *Shewanella putrefaciens* (*S. putrefaciens*) TadA, *Haemophilus influenzae* F3031 (*H. influenzae*) TadA, *Caulobacter crescentus* (*C. crescentus*) TadA, *Geobacter sulfurreducens* (*G. sulfurreducens*) TadA, or TadA*7.10.

In some embodiments, an adenosine deaminase is a TadA*8. In one embodiment, an adenosine deaminase is a TadA*8 that comprises or consists essentially of the following sequence or a fragment thereof having adenosine deaminase activity:
MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 404)

In some embodiments, the TadA*8 is truncated. In some embodiments, the truncated TadA*8 is missing 1, 2, 3, 4, 5 ,6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 6, 17, 18, 19, or 20 N-terminal amino acid residues relative to the full length TadA*8. In some embodiments, the truncated TadA*8 is missing 1, 2, 3, 4, 5 ,6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 6, 17, 18, 19, or 20 C-terminal amino acid residues relative to the full length TadA*8. In some embodiments the adenosine deaminase variant is a full-length TadA*8.

In some embodiments the TadA*8 is TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, TadA*8.7, TadA*8.8, TadA*8.9, TadA*8.10, TadA*8.11, TadA*8.12, TadA*8.13, TadA*8.14, TadA*8.15, TadA*8.16, TadA*8.17, TadA*8.18, TadA*8.19, TadA*8.20, TadA*8.21, TadA*8.22, TadA*8.23, or TadA*8.24.

In other embodiments, a base editor of the disclosure comprising an adenosine deaminase variant (*e.g.*, TadA*8) monomer comprising one or more of the following alterations: R26C, V88A, A109S, T111R, D119N, H122N, Y147D, F149Y, T166I and/or D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant (TadA*8) monomer comprises a combination of alterations selected from the group of: R26C + A109S + T111R

+ D119N + H122N + Y147D + F149Y + T166I + D167N; V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; V88A + T111R + D119N + F149Y; and A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, a base editor comprises a heterodimer of a wild-type adenosine deaminase domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising one or more of the following alterations R26C, V88A, A109S, T111R, D119N, H122N, Y147D, F149Y, T166I and/or D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the base editor comprises a heterodimer of a wild-type adenosine deaminase domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising a combination of alterations selected from the group of: R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N; V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; V88A + T111R + D119N + F149Y; and A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, a base editor comprises a heterodimer of a TadA*7.10 domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising one or more of the following alterations R26C, V88A, A109S, T111R, D119N, H122N, Y147D, F149Y, T166I and/or D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the base editor comprises a heterodimer of a TadA*7.10 domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising a combination of alterations selected from the group of: R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N; V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; V88A + T111R + D119N + F149Y; and A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, the adenosine deaminase variant is a heterodimer of a TadA*7.10 domain and an adenosine deaminase variant domain (*e.g.*, TadA*7.10) comprising one or more of the following alterations L36H, I76Y, V82G, Y147T, Y147D, F149Y, Q154S, N157K, and/or D167N, relative to TadA*7.10, the TadA reference sequence,

or a corresponding mutation in another TadA. In some embodiments, an adenosine deaminase variant is a heterodimer comprising a TadA*7.10 domain and an adenosine deaminase variant domain (*e.g.*, MSP828) having the following alterations V82G, Y147T/D, Q154S, and one or more of L36H, I76Y, F149Y, N157K, and D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a heterodimer of a TadA*7.10 domain and an adenosine deaminase variant domain (*e.g.*, TadA*7.10) comprising a combination of alterations selected from the group of: V82G + Y147T + Q154S; I76Y + V82G + Y147T + Q154S; L36H + V82G + Y147T + Q154S + N157K; V82G + Y147D + F149Y + Q154S + D167N; L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N; L36H + I76Y + V82G + Y147T + Q154S + N157K; I76Y + V82G + Y147D + F149Y + Q154S + D167N; L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In some embodiments, the TadA*8 is a variant as shown in Table 5. Table 5 shows certain amino acid position numbers in the TadA amino acid sequence and the amino acids present in those positions in the TadA-7.10 adenosine deaminase. Table 5 also shows amino acid changes in TadA variants relative to TadA-7.10 following phage-assisted non-continuous evolution (PANCE) and phage-assisted continuous evolution (PACE), as described in M. Richter *et al.*, 2020, Nature Biotechnology, doi.org/10.1038/s41587-020-0453-z, the entire contents of which are incorporated by reference herein. In some embodiments, the TadA*8 is TadA*8a, TadA*8b, TadA*8c, TadA*8d, or TadA*8e. In some embodiments, the TadA*8 is TadA*8e.

Table 5. Select TadA*8 Variants

		TadA amino acid number									
	TadA	26	88	109	111	119	122	147	149	166	167
	TadA-7.10	R	V	A	T	D	H	Y	F	T	D
PANCE 1					R						
PANCE 2				S/T	R						
	TadA-8a	C		S	R	N	N	D	Y	I	N
	TadA-8b		A	S	R	N	N		Y	I	N
PACE	TadA-8c	C		S	R	N	N		Y	I	N
	TadA-8d		A		R	N			Y		
	TadA-8e			S	R	N	N	D	Y	I	N

In some embodiments, the TadA variant is a variant as shown in **Table 5.1**. **Table 5.1** shows certain amino acid position numbers in the TadA amino acid sequence and the amino

acids present in those positions in the TadA*7.10 adenosine deaminase. In some embodiments, the TadA variant is MSP605, MSP680, MSP823, MSP824, MSP825, MSP827, MSP828, or MSP829. In some embodiments, the TadA variant is MSP828. In some embodiments, the TadA variant is MSP829.

Table 5.1. TadA Variants

Variant	TadA Amino Acid Number							
	36	**76**	**82**	**147**	**149**	**154**	**157**	**167**
TadA-7.10	L	I	V	Y	F	Q	N	D
MSP605			G	T		S		
MSP680		Y	G	T		S		
MSP823	H		G	T		S	K	
MSP824			G	D	Y	S		N
MSP825	H		G	D	Y	S	K	N
MSP827	H	Y	G	T		S	K	
MSP828		Y	G	D	Y	S		N
MSP829	H	Y	G	D	Y	S	K	N

In one embodiment, a fusion protein of the invention comprises a wild-type TadA is linked to an adenosine deaminase variant described herein (*e.g.*, TadA*8), which is linked to Cas9 nickase. In particular embodiments, the fusion proteins comprise a single TadA*8 domain (*e.g.*, provided as a monomer). In other embodiments, the fusion protein comprises TadA*8 and TadA(wt), which are capable of forming heterodimers.

In some embodiments, the adenosine deaminase comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of the amino acid sequences set forth in any of the adenosine deaminases provided herein. It should be appreciated that adenosine deaminases provided herein may include one or more mutations (*e.g.*, any of the mutations provided herein). The disclosure provides any deaminase domains with a certain percent identity plus any of the mutations or combinations thereof described herein. In some embodiments, the adenosine deaminase comprises an amino acid sequence that has 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, or more mutations compared to a reference sequence, or any of the adenosine deaminases provided herein. In some embodiments, the adenosine deaminase comprises an amino acid sequence that has at least 5, at least 10, at least

15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 110, at least 120, at least 130, at least 140, at least 150, at least 160, or at least 170 identical contiguous amino acid residues as compared to any one of the amino acid sequences known in the art or described herein.

In particular embodiments, a TadA*8 comprises one or more mutations at any of the following positions shown in bold. In other embodiments, a TadA*8 comprises one or more mutations at any of the positions shown with underlining:

MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG [50]
LHDPTAHAEI MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG [100]
RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCYFFR [150]
MPRQVFNAQK KAQSSTD (SEQ ID NO: 1)

For example, the TadA*8 comprises alterations at amino acid position 82 and/or 166 (*e.g.*, V82S, T166R) alone or in combination with any one or more of the following Y147T, Y147R, Q154S, Y123H, and/or Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In particular embodiments, a combination of alterations is selected from the group of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R +Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In some embodiments, the TadA*8 is truncated. In some embodiments, the truncated TadA*8 is missing 1, 2, 3, 4, 5 ,6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 6, 17, 18, 19, or 20 N-terminal amino acid residues relative to the full length TadA*8. In some embodiments, the truncated TadA*8 is missing 1, 2, 3, 4, 5 ,6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 6, 17, 18, 19, or 20 C-terminal amino acid residues relative to the full length TadA*8. In some embodiments the adenosine deaminase variant is a full-length TadA*8.

In one embodiment, a fusion protein of the invention comprises a wild-type TadA is linked to an adenosine deaminase variant described herein (*e.g.*, TadA*8), which is linked to Cas9 nickase. In particular embodiments, the fusion proteins comprise a single TadA*8 domain (*e.g.*, provided as a monomer). In other embodiments, the base editor comprises TadA*8 and TadA(wt), which are capable of forming heterodimers.

In particular embodiments, the fusion proteins comprise a single (*e.g.*, provided as a monomer) TadA*8. In some embodiments, the TadA*8 is linked to a Cas9 nickase. In some embodiments, the fusion proteins of the invention comprise as a heterodimer of a wild-type TadA (TadA(wt)) linked to a TadA*8. In other embodiments, the fusion proteins of the invention comprise as a heterodimer of a TadA*7.10 linked to a TadA*8. In some embodiments, the base editor is ABE8 comprising a TadA*8 variant monomer. In some embodiments, the base editor is ABE8 comprising a heterodimer of a TadA*8 and a TadA(wt). In some embodiments, the base editor is ABE8 comprising a heterodimer of a TadA*8 and TadA*7.10. In some embodiments, the base editor is ABE8 comprising a heterodimer of a TadA*8. In some embodiments, the TadA*8 is selected from Table 5, 11 or 12. In some embodiments, the ABE8 is selected from Table 11, 12 or 14.

In some embodiments, the adenosine deaminase is a TadA*9 variant. In some embodiments, the adenosine deaminase is a TadA*9 variant selected from the variants described below and with reference to the following sequence (termed TadA*7.10): MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG LHDPTAHAEI MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCYFFR MPRQVFNAQK KAQSSTD (SEQ ID NO: 1)

In some embodiments, an adenosine deaminase comprises one or more of the following alterations: R21N, R23H, E25F, N38G, L51W, P54C, M70V, Q71M, N72K, Y73S, V82T, M94V, P124W, T133K, D139L, D139M, C146R, and A158K. The one or more alternations are shown in the sequence above in underlining and bold font.

In some embodiments, an adenosine deaminase comprises one or more of the following combinations of alterations: V82S + Q154R + Y147R; V82S + Q154R + Y123H; V82S + Q154R + Y147R+ Y123H; Q154R + Y147R + Y123H + I76Y+ V82S; V82S + I76Y; V82S + Y147R; V82S + Y147R + Y123H; V82S + Q154R + Y123H; Q154R + Y147R + Y123H + I76Y; V82S + Y147R; V82S + Y147R + Y123H; V82S + Q154R + Y123H; V82S + Q154R + Y147R; V82S + Q154R + Y147R; Q154R + Y147R + Y123H + I76Y; Q154R + Y147R + Y123H + I76Y + V82S; I76Y_V82S_Y123H_Y147R_Q154R; Y147R + Q154R + H123H; and V82S + Q154R.

In some embodiments, an adenosine deaminase comprises one or more of the following combinations of alterations: E25F + V82S + Y123H, T133K + Y147R + Q154R; E25F + V82S + Y123H + Y147R + Q154R; L51W + V82S + Y123H + C146R + Y147R + Q154R; Y73S + V82S + Y123H + Y147R + Q154R; P54C + V82S + Y123H + Y147R +

Q154R; N38G + V82T + Y123H + Y147R + Q154R; N72K + V82S + Y123H + D139L + Y147R + Q154R; E25F + V82S + Y123H + D139M + Y147R + Q154R; Q71M + V82S + Y123H + Y147R + Q154R; E25F + V82S + Y123H + T133K + Y147R + Q154R; E25F + V82S + Y123H + Y147R + Q154R; V82S + Y123H + P124W + Y147R + Q154R; L51W + V82S + Y123H + C146R + Y147R + Q154R; P54C + V82S + Y123H + Y147R + Q154R; Y73S + V82S + Y123H + Y147R + Q154R; N38G + V82T + Y123H + Y147R + Q154R; R23H + V82S + Y123H + Y147R + Q154R; R21N + V82S + Y123H + Y147R + Q154R; V82S + Y123H + Y147R + Q154R + A158K; N72K + V82S + Y123H + D139L + Y147R + Q154R; E25F + V82S + Y123H + D139M + Y147R + Q154R; and M70V + V82S + M94V + Y123H + Y147R + Q154R

In some embodiments, an adenosine deaminase comprises one or more of the following combinations of alterations: Q71M + V82S + Y123H + Y147R + Q154R; E25F + I76Y+ V82S + Y123H + Y147R + Q154R; I76Y + V82T + Y123H + Y147R + Q154R; N38G + I76Y + V82S + Y123H + Y147R + Q154R; R23H + I76Y + V82S + Y123H + Y147R + Q154R; P54C + I76Y + V82S + Y123H + Y147R + Q154R; R21N + I76Y + V82S + Y123H + Y147R + Q154R; I76Y + V82S + Y123H + D139M + Y147R + Q154R; Y73S + I76Y + V82S + Y123H + Y147R + Q154R; E25F + I76Y + V82S + Y123H + Y147R + Q154R; I76Y + V82T + Y123H + Y147R + Q154R; N38G + I76Y + V82S + Y123H + Y147R + Q154R; R23H + I76Y + V82S + Y123H + Y147R + Q154R; P54C + I76Y + V82S + Y123H + Y147R + Q154R; R21N + I76Y + V82S + Y123H + Y147R + Q154R; I76Y + V82S + Y123H + D139M + Y147R + Q154R; Y73S + I76Y + V82S + Y123H + Y147R + Q154R; and V82S + Q154R; N72K_V82S + Y123H + Y147R + Q154R; Q71M_V82S + Y123H + Y147R + Q154R; V82S + Y123H + T133K + Y147R + Q154R; V82S + Y123H + T133K + Y147R + Q154R + A158K; M70V +Q71M +N72K +V82S + Y123H + Y147R + Q154R; N72K_V82S + Y123H + Y147R + Q154R; Q71M_V82S + Y123H + Y147R + Q154R; M70V +V82S + M94V + Y123H + Y147R + Q154R; V82S + Y123H + T133K + Y147R + Q154R; V82S + Y123H + T133K + Y147R + Q154R + A158K; and M70V +Q71M +N72K +V82S + Y123H + Y147R + Q154R. In some embodiments, the adenosine deaminase is expressed as a monomer. In other embodiments, the adenosine deaminase is expressed as a heterodimer. In some embodiments, the deaminase or other polypeptide sequence lacks a methionine, for example when included as a component of a fusion protein. This can alter the numbering of positions. However, the skilled person will understand that such corresponding mutations refer to the same mutation, *e.g.*, Y73S and Y72S and D139M and D138M.

In some embodiments, the TadA*9 variant comprises the alterations described in Table 15 as described herein. In some embodiments, the TadA*9 variant is a monomer. In some embodiments, the TadA*9 variant is a heterodimer with a wild-type TadA adenosine deaminase. In some embodiments, the TadA*9 variant is a heterodimer with another TadA variant (*e.g.*, TadA*8, TadA*9). Additional details of TadA*9 adenosine deaminases are described in International PCT Application No. PCT/2020/049975, which is incorporated herein by reference for its entirety.

Any of the mutations provided herein and any additional mutations (*e.g.*, based on the ecTadA amino acid sequence) can be introduced into any other adenosine deaminases. Any of the mutations provided herein can be made individually or in any combination in TadA reference sequence or another adenosine deaminase (*e.g.*, ecTadA).

Details of A to G nucleobase editing proteins are described in International PCT Application No. PCT/2017/045381 (WO2018/027078) and Gaudelli, N.M., *et al.*, "Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage" Nature, 551, 464-471 (2017), the entire contents of which are hereby incorporated by reference.

C to T Editing

In some embodiments, a base editor disclosed herein comprises a fusion protein comprising cytidine deaminase capable of deaminating a target cytidine (C) base of a polynucleotide to produce uridine (U), which has the base pairing properties of thymine. In some embodiments, for example where the polynucleotide is double-stranded (*e.g.*, DNA), the uridine base can then be substituted with a thymidine base (*e.g.*, by cellular repair machinery) to give rise to a C:G to a T:A transition. In other embodiments, deamination of a C to U in a nucleic acid by a base editor cannot be accompanied by substitution of the U to a T.

The deamination of a target C in a polynucleotide to give rise to a U is a non-limiting example of a type of base editing that can be executed by a base editor described herein. In another example, a base editor comprising a cytidine deaminase domain can mediate conversion of a cytosine (C) base to a guanine (G) base. For example, a U of a polynucleotide produced by deamination of a cytidine by a cytidine deaminase domain of a base editor can be excised from the polynucleotide by a base excision repair mechanism (*e.g.*, by a uracil DNA glycosylase (UDG) domain), producing an abasic site. The nucleobase opposite the abasic site can then be substituted (*e.g.*, by base repair machinery) with another

base, such as a C, by for example a translesion polymerase. Although it is typical for a nucleobase opposite an abasic site to be replaced with a C, other substitutions (*e.g.*, A, G or T) can also occur.

Accordingly, in some embodiments a base editor described herein comprises a deamination domain (*e.g.*, cytidine deaminase domain) capable of deaminating a target C to a U in a polynucleotide. Further, as described below, the base editor can comprise additional domains which facilitate conversion of the U resulting from deamination to, in some embodiments, a T or a G. For example, a base editor comprising a cytidine deaminase domain can further comprise a uracil glycosylase inhibitor (UGI) domain to mediate substitution of a U by a T, completing a C-to-T base editing event. In another example, a base editor can incorporate a translesion polymerase to improve the efficiency of C-to-G base editing, since a translesion polymerase can facilitate incorporation of a C opposite an abasic site (*i.e.*, resulting in incorporation of a G at the abasic site, completing the C-to-G base editing event).

A base editor comprising a cytidine deaminase as a domain can deaminate a target C in any polynucleotide, including DNA, RNA and DNA-RNA hybrids. Typically, a cytidine deaminase catalyzes a C nucleobase that is positioned in the context of a single-stranded portion of a polynucleotide. In some embodiments, the entire polynucleotide comprising a target C can be single-stranded. For example, a cytidine deaminase incorporated into the base editor can deaminate a target C in a single-stranded RNA polynucleotide. In other embodiments, a base editor comprising a cytidine deaminase domain can act on a double-stranded polynucleotide, but the target C can be positioned in a portion of the polynucleotide which at the time of the deamination reaction is in a single-stranded state. For example, in embodiments where the NAGPB domain comprises a Cas9 domain, several nucleotides can be left unpaired during formation of the Cas9-gRNA-target DNA complex, resulting in formation of a Cas9 "R-loop complex". These unpaired nucleotides can form a bubble of single-stranded DNA that can serve as a substrate for a single-strand specific nucleotide deaminase enzyme (*e.g.*, cytidine deaminase).

In some embodiments, a cytidine deaminase of a base editor can comprise all or a portion of an apolipoprotein B mRNA editing complex (APOBEC) family deaminase. APOBEC is a family of evolutionarily conserved cytidine deaminases. Members of this family are C-to-U editing enzymes. The N-terminal domain of APOBEC like proteins is the catalytic domain, while the C-terminal domain is a pseudocatalytic domain. More specifically, the catalytic domain is a zinc dependent cytidine deaminase domain and is

important for cytidine deamination. APOBEC family members include APOBEC1, APOBEC2, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D ("APOBEC3E" now refers to this), APOBEC3F, APOBEC3G, APOBEC3H, APOBEC4, and Activation-induced (cytidine) deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of an APOBEC1 deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of APOBEC2 deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of is an APOBEC3 deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of an APOBEC3A deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of APOBEC3B deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of APOBEC3C deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of APOBEC3D deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of APOBEC3E deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of APOBEC3F deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of APOBEC3G deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of APOBEC3H deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of APOBEC4 deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of activation-induced deaminase (AID). In some embodiments a deaminase incorporated into a base editor comprises all or a portion of cytidine deaminase 1 (CDA1). It should be appreciated that a base editor can comprise a deaminase from any suitable organism (*e.g.*, a human or a rat). In some embodiments, a deaminase domain of a base editor is from a human, chimpanzee, gorilla, monkey, cow, dog, rat, or mouse. In some embodiments, the deaminase domain of the base editor is derived from rat (*e.g.*, rat APOBEC1). In some embodiments, the deaminase domain of the base editor is human APOBEC1. In some embodiments, the deaminase domain of the base editor is pmCDA1.

Other exemplary deaminases that can be fused to Cas9 according to aspects of this disclosure are provided below. In embodiments, the deaminases are activation-induced deaminases (AID). It should be understood that, in some embodiments, the active domain of the respective sequence can be used, *e.g.*, the domain without a localizing signal (nuclear localization sequence, without nuclear export signal, cytoplasmic localizing signal).

Some aspects of the present disclosure are based on the recognition that modulating the deaminase domain catalytic activity of any of the fusion proteins described herein, for example by making point mutations in the deaminase domain, affect the processivity of the fusion proteins (*e.g.*, base editors). For example, mutations that reduce, but do not eliminate, the catalytic activity of a deaminase domain within a base editing fusion protein can make it less likely that the deaminase domain will catalyze the deamination of a residue adjacent to a target residue, thereby narrowing the deamination window. The ability to narrow the deamination window can prevent unwanted deamination of residues adjacent to specific target residues, which can decrease or prevent off-target effects.

For example, in some embodiments, an APOBEC deaminase incorporated into a base editor can comprise one or more mutations selected from the group consisting of H121X, H122X, R126X, R126X, R118X, W90X, W90X, and R132X of rAPOBEC1, or one or more corresponding mutations in another APOBEC deaminase, wherein X is any amino acid. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise one or more mutations selected from the group consisting of H121R, H122R, R126A, R126E, R118A, W90A, W90Y, and R132E of rAPOBEC1, or one or more corresponding mutations in another APOBEC deaminase.

In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise one or more mutations selected from the group consisting of D316X, D317X, R320X, R320X, R313X, W285X, W285X, R326X of hAPOBEC3G, or one or more corresponding mutations in another APOBEC deaminase, wherein X is any amino acid. In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising one or more mutations selected from the group consisting of D316R, D317R, R320A, R320E, R313A, W285A, W285Y, R326E of hAPOBEC3G, or one or more corresponding mutations in another APOBEC deaminase.

In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise a H121R and a H122R mutation of rAPOBEC1, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a R126A mutation of rAPOBEC1, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a R126E mutation of rAPOBEC1, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase

comprising a R118A mutation of rAPOBEC1, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a W90A mutation of rAPOBEC1, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a W90Y mutation of rAPOBEC1, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a R132E mutation of rAPOBEC1, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a W90Y and a R126E mutation of rAPOBEC1, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a R126E and a R132E mutation of rAPOBEC1, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a W90Y and a R132E mutation of rAPOBEC1, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a W90Y, R126E, and R132E mutation of rAPOBEC1, or one or more corresponding mutations in another APOBEC deaminase.

In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a D316R and a D317R mutation of hAPOBEC3G, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising a R320A mutation of hAPOBEC3G, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a R320E mutation of hAPOBEC3G, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a R313A mutation of hAPOBEC3G, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a W285A mutation of hAPOBEC3G, or one or more corresponding mutations in

another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a W285Y mutation of hAPOBEC3G, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a R326E mutation of hAPOBEC3G, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a W285Y and a R320E mutation of hAPOBEC3G, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a R320E and a R326E mutation of hAPOBEC3G, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a W285Y and a R326E mutation of hAPOBEC3G, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a W285Y, R320E, and R326E mutation of hAPOBEC3G, or one or more corresponding mutations in another APOBEC deaminase.

A number of modified cytidine deaminases are commercially available, including, but not limited to, SaBE3, SaKKH-BE3, VQR-BE3, EQR-BE3, VRER-BE3, YE1-BE3, EE-BE3, YE2-BE3, and YEE-BE3, which are available from Addgene (plasmids 85169, 85170, 85171, 85172, 85173, 85174, 85175, 85176, 85177). In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of an APOBEC1 deaminase.

In some embodiments, the fusion proteins of the invention comprise one or more cytidine deaminase domains. In some embodiments, the cytidine deaminases provided herein are capable of deaminating cytosine or 5-methylcytosine to uracil or thymine. In some embodiments, the cytidine deaminases provided herein are capable of deaminating cytosine in DNA. The cytidine deaminase may be derived from any suitable organism. In some embodiments, the cytidine deaminase is a naturally-occurring cytidine deaminase that includes one or more mutations corresponding to any of the mutations provided herein. One of skill in the art will be able to identify the corresponding residue in any homologous protein, *e.g.*, by sequence alignment and determination of homologous residues. Accordingly, one of skill in the art would be able to generate mutations in any naturally-occurring cytidine deaminase that corresponds to any of the mutations described herein. In some embodiments, the cytidine deaminase is from a prokaryote. In some embodiments, the cytidine deaminase is

from a bacterium. In some embodiments, the cytidine deaminase is from a mammal (*e.g.*, human).

In some embodiments, the cytidine deaminase comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of the cytidine deaminase amino acid sequences set forth herein. It should be appreciated that cytidine deaminases provided herein may include one or more mutations (*e.g.*, any of the mutations provided herein). Some embodiments provide a polynucleotide molecule encoding the cytidine deaminase nucleobase editor polypeptide of any previous aspect or as delineated herein. In some embodiments, the polynucleotide is codon optimized.

The disclosure provides any deaminase domains with a certain percent identity plus any of the mutations or combinations thereof described herein. In some embodiments, the cytidine deaminase comprises an amino acid sequence that has 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, or more mutations compared to a reference sequence, or any of the cytidine deaminases provided herein. In some embodiments, the cytidine deaminase comprises an amino acid sequence that has at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 110, at least 120, at least 130, at least 140, at least 150, at least 160, or at least 170 identical contiguous amino acid residues as compared to any one of the amino acid sequences known in the art or described herein.

A fusion protein of the invention second protein comprises two or more nucleic acid editing domains.

Details of C to T nucleobase editing proteins are described in International PCT Application No. PCT/US2016/058344 (WO2017/070632) and Komor, A.C., *et al.*, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016), the entire contents of which are hereby incorporated by reference.

Guide Polynucleotides

A polynucleotide programmable nucleotide binding domain, when in conjunction with a bound guide polynucleotide (*e.g.*, gRNA), can specifically bind to a target polynucleotide sequence (*i.e.*, via complementary base pairing between bases of the bound guide nucleic acid and bases of the target polynucleotide sequence) and thereby localize the base editor to the target nucleic acid sequence desired to be edited. In some embodiments, the target polynucleotide sequence comprises single-stranded DNA or double-stranded DNA. In some embodiments, the target polynucleotide sequence comprises RNA. In some embodiments, the target polynucleotide sequence comprises a DNA-RNA hybrid.

CRISPR is an adaptive immune system that provides protection against mobile genetic elements (viruses, transposable elements and conjugative plasmids). CRISPR clusters contain spacers, sequences complementary to antecedent mobile elements, and target invading nucleic acids. CRISPR clusters are transcribed and processed into CRISPR RNA (crRNA). In type II CRISPR systems, correct processing of pre-crRNA requires a trans-encoded small RNA (tracrRNA), endogenous ribonuclease 3 (rnc) and a Cas9 protein. The tracrRNA serves as a guide for ribonuclease 3-aided processing of pre-crRNA. Subsequently, Cas9/crRNA/tracrRNA endonucleolytically cleaves linear or circular dsDNA target complementary to the spacer. The target strand not complementary to crRNA is first cut endonucleolytically, and then trimmed 3′-5′ exonucleolytically. In nature, DNA-binding and cleavage typically requires protein and both RNAs. However, single guide RNAs ("sgRNA", or simply "gRNA") can be engineered so as to incorporate aspects of both the crRNA and tracrRNA into a single RNA species. *See, e.g.*, Jinek M., *et al.* Science 337:816-821(2012), the entire contents of which is hereby incorporated by reference. Cas9 recognizes a short motif in the CRISPR repeat sequences (the PAM or protospacer adjacent motif) to help distinguish self versus non-self. See *e.g.*, "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*." Ferretti, J.J. *et al.*, Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E. *et al.*, Nature 471:602-607(2011); and "Programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M.*et al*, Science 337:816-821(2012), the entire contents of each of which are incorporated herein by reference).

The PAM sequence can be any PAM sequence known in the art. Suitable PAM sequences include, but are not limited to, NGG, NGA, NGC, NGN, NGT, NGCG, NGAG, NGAN, NGNG, NGCN, NGCG, NGTN, NNGRRT, NNNRRT, NNGRR(N), TTTV, TYCV,

TYCV, TATV, NNNNGATT, NNAGAAW, or NAAAAC. Y is a pyrimidine; N is any nucleotide base; W is A or T.

In an embodiment, a guide polynucleotide described herein can be RNA or DNA. In one embodiment, the guide polynucleotide is a gRNA. An RNA/Cas complex can assist in "guiding" a Cas protein to a target DNA. Cas9/crRNA/tracrRNA endonucleolytically cleaves linear or circular dsDNA target complementary to the spacer. The target strand not complementary to crRNA is first cut endonucleolytically, then trimmed 3'-5' exonucleolytically. In nature, DNA-binding and cleavage typically requires protein and both RNAs. However, single guide RNAs ("sgRNA", or simply "gRNA") can be engineered so as to incorporate aspects of both the crRNA and tracrRNA into a single RNA species. *See, e.g.*, Jinek M. *et al.*, Science 337:816-821(2012), the entire contents of which is hereby incorporated by reference.

In some embodiments, the guide polynucleotide is at least one single guide RNA ("sgRNA" or "gRNA"). In some embodiments, a guide polynucleotide comprises two or more individual polynucleotides, which can interact with one another via for example complementary base pairing (*e.g.*, a dual guide polynucleotide, dual gRNA). For example, a guide polynucleotide can comprise a CRISPR RNA (crRNA) and a trans-activating CRISPR RNA (tracrRNA) or can comprise one or more trans-activating CRISPR RNA (tracrRNA).

In some embodiments, the guide polynucleotide is at least one tracrRNA. In some embodiments, the guide polynucleotide does not require PAM sequence to guide the polynucleotide-programmable DNA-binding domain (*e.g.*, Cas9 or Cpf1) to the target nucleotide sequence.

A guide polynucleotide may include natural or non-natural (or unnatural) nucleotides (*e.g.*, peptide nucleic acid or nucleotide analogs). In some cases, the targeting region of a guide nucleic acid sequence can be at least 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides in length. A targeting region of a guide nucleic acid can be between 10-30 nucleotides in length, or between 15-25 nucleotides in length, or between 15-20 nucleotides in length.

In some embodiments, the base editor provided herein utilizes one or more guide polynucleotide (*e.g.*, multiple gRNA). In some embodiments, a single guide polynucleotide is utilized for different base editors described herein. For example, a single guide polynucleotide can be utilized for a cytidine base editor and an adenosine base editor.

In some embodiments, the methods described herein can utilize an engineered Cas protein. A guide RNA (gRNA) is a short synthetic RNA composed of a scaffold sequence

necessary for Cas-binding and a user-defined ~20 nucleotide spacer that defines the genomic target to be modified. Exemplary gRNA scaffold sequences are provided in the sequence listing as SEQ ID NOs: 405-415. Thus, a skilled artisan can change the genomic target of the Cas protein specificity is partially determined by how specific the gRNA targeting sequence is for the genomic target compared to the rest of the genome.

In other embodiments, a guide polynucleotide can comprise both the polynucleotide targeting portion of the nucleic acid and the scaffold portion of the nucleic acid in a single molecule (*i.e.*, a single-molecule guide nucleic acid). For example, a single-molecule guide polynucleotide can be a single guide RNA (sgRNA or gRNA). Herein the term guide polynucleotide sequence contemplates any single, dual or multi-molecule nucleic acid capable of interacting with and directing a base editor to a target polynucleotide sequence.

Typically, a guide polynucleotide (*e.g.*, crRNA/trRNA complex or a gRNA) comprises a "polynucleotide-targeting segment" that includes a sequence capable of recognizing and binding to a target polynucleotide sequence, and a "protein-binding segment" that stabilizes the guide polynucleotide within a polynucleotide programmable nucleotide binding domain component of a base editor. In some embodiments, the polynucleotide targeting segment of the guide polynucleotide recognizes and binds to a DNA polynucleotide, thereby facilitating the editing of a base in DNA. In other cases, the polynucleotide targeting segment of the guide polynucleotide recognizes and binds to an RNA polynucleotide, thereby facilitating the editing of a base in RNA. Herein a "segment" refers to a section or region of a molecule, *e.g.*, a contiguous stretch of nucleotides in the guide polynucleotide. A segment can also refer to a region/section of a complex such that a segment can comprise regions of more than one molecule. For example, where a guide polynucleotide comprises multiple nucleic acid molecules, the protein-binding segment of can include all or a portion of multiple separate molecules that are for instance hybridized along a region of complementarity. In some embodiments, a protein-binding segment of a DNA-targeting RNA that comprises two separate molecules can comprise (i) base pairs 40-75 of a first RNA molecule that is 100 base pairs in length; and (ii) base pairs 10-25 of a second RNA molecule that is 50 base pairs in length. The definition of "segment," unless otherwise specifically defined in a particular context, is not limited to a specific number of total base pairs, is not limited to any particular number of base pairs from a given RNA molecule, is not limited to a particular number of separate molecules within a complex, and can include regions of RNA molecules that are of any total length and can include regions with complementarity to other molecules.

The guide polynucleotides can be synthesized chemically, synthesized enzymatically, or a combination thereof. For example, the gRNA can be synthesized using standard phosphoramidite-based solid-phase synthesis methods. Alternatively, the gRNA can be synthesized *in vitro* by operably linking DNA encoding the gRNA to a promoter control sequence that is recognized by a phage RNA polymerase. Examples of suitable phage promoter sequences include T7, T3, SP6 promoter sequences, or variations thereof. In embodiments in which the gRNA comprises two separate molecules (*e.g.*., crRNA and tracrRNA), the crRNA can be chemically synthesized and the tracrRNA can be enzymatically synthesized.

A guide polynucleotide may be expressed, for example, by a DNA that encodes the gRNA, *e.g.*, a DNA vector comprising a sequence encoding the gRNA. The gRNA may be encoded alone or together with an encoded base editor. Such DNA sequences may be introduced into an expression system, *e.g.*, a cell, together or separately. For example, DNA sequences encoding a polynucleotide programmable nucleotide binding domain and a gRNA may be introduced into a cell, each DNA sequence can be part of a separate molecule (*e.g.*, one vector containing the polynucleotide programmable nucleotide binding domain coding sequence and a second vector containing the gRNA coding sequence) or both can be part of a same molecule (*e.g.*, one vector containing coding (and regulatory) sequence for both the polynucleotide programmable nucleotide binding domain and the gRNA). An RNA can be transcribed from a synthetic DNA molecule, *e.g.*, a gBlocks® gene fragment. A gRNA molecule can be transcribed *in vitro*.

A gRNA or a guide polynucleotide can comprise three regions: a first region at the 5' end that can be complementary to a target site in a chromosomal sequence, a second internal region that can form a stem loop structure, and a third 3' region that can be single-stranded. A first region of each gRNA can also be different such that each gRNA guides a fusion protein to a specific target site. Further, second and third regions of each gRNA can be identical in all gRNAs.

A first region of a gRNA or a guide polynucleotide can be complementary to sequence at a target site in a chromosomal sequence such that the first region of the gRNA can base pair with the target site. In some cases, a first region of a gRNA can comprise from or from about 10 nucleotides to 25 nucleotides (*i.e.*, from 10 nucleotides to nucleotides; or from about 10 nucleotides to about 25 nucleotides; or from 10 nucleotides to about 25 nucleotides; or from about 10 nucleotides to 25 nucleotides) or more. For example, a region of base pairing between a first region of a gRNA and a target site in a chromosomal sequence

can be or can be about 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25, or more nucleotides in length. Sometimes, a first region of a gRNA can be or can be about 19, 20, or 21 nucleotides in length.

A gRNA or a guide polynucleotide can also comprise a second region that forms a secondary structure. For example, a secondary structure formed by a gRNA can comprise a stem (or hairpin) and a loop. A length of a loop and a stem can vary. For example, a loop can range from or from about 3 to 10 nucleotides in length, and a stem can range from or from about 6 to 20 base pairs in length. A stem can comprise one or more bulges of 1 to 10 or about 10 nucleotides. The overall length of a second region can range from or from about 16 to 60 nucleotides in length. For example, a loop can be or can be about 4 nucleotides in length and a stem can be or can be about 12 base pairs.

A gRNA or a guide polynucleotide can also comprise a third region at the 3' end that can be essentially single-stranded. For example, a third region is sometimes not complementarity to any chromosomal sequence in a cell of interest and is sometimes not complementarity to the rest of a gRNA. Further, the length of a third region can vary. A third region can be more than or more than about 4 nucleotides in length. For example, the length of a third region can range from or from about 5 to 60 nucleotides in length.

A gRNA or a guide polynucleotide can target any exon or intron of a gene target. In some cases, a guide can target exon 1 or 2 of a gene, in other cases; a guide can target exon 3 or 4 of a gene. In some embodiments, a composition comprises multiple gRNAs that all target the same exon or multiple gRNAs that target different exons. An exon and/or an intron of a gene can be targeted.

A gRNA or a guide polynucleotide can target a nucleic acid sequence of about 20 nucleotides or less than about 20 nucleotides (*e.g.*, at least about 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30 nucleotides), or anywhere between about 1-100 nucleotides (*e.g.*, 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 40, 50, 60, 70, 80, 90, 100). A target nucleic acid sequence can be or can be about 20 bases immediately 5' of the first nucleotide of the PAM. A gRNA can target a nucleic acid sequence. A target nucleic acid can be at least or at least about 1-10, 1-20, 1-30, 1-40, 1-50, 1-60, 1-70, 1-80, 1-90, or 1-100 nucleotides.

Methods for selecting, designing, and validating guide polynucleotides, *e.g.*, gRNAs and targeting sequences are described herein and known to those skilled in the art. For example, to minimize the impact of potential substrate promiscuity of a deaminase domain in the nucleobase editor system (*e.g.*, an AID domain), the number of residues that could unintentionally be targeted for deamination (*e.g.*, off-target C residues that could potentially

reside on single strand DNA within the target nucleic acid locus) may be minimized. In addition, software tools can be used to optimize the gRNAs corresponding to a target nucleic acid sequence, *e.g.*, to minimize total off-target activity across the genome. For example, for each possible targeting domain choice using *S. pyogenes* Cas9, all off-target sequences (preceding selected PAMs, *e.g.*, NAG or NGG) may be identified across the genome that contain up to certain number (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) of mismatched base-pairs. First regions of gRNAs complementary to a target site can be identified, and all first regions (*e.g.*, crRNAs) can be ranked according to its total predicted off-target score; the top-ranked targeting domains represent those that are likely to have the greatest on-target and the least off-target activity. Candidate targeting gRNAs can be functionally evaluated by using methods known in the art and/or as set forth herein.

As a non-limiting example, target DNA hybridizing sequences in crRNAs of a gRNA for use with Cas9s may be identified using a DNA sequence searching algorithm. gRNA design is carried out using custom gRNA design software based on the public tool cas-OFFinder as described in Bae S., Park J., & Kim J.-S. Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics 30, 1473-1475 (2014). This software scores guides after calculating their genome-wide off-target propensity. Typically matches ranging from perfect matches to 7 mismatches are considered for guides ranging in length from 17 to 24. Once the off-target sites are computationally-determined, an aggregate score is calculated for each guide and summarized in a tabular output using a web-interface. In addition to identifying potential target sites adjacent to PAM sequences, the software also identifies all PAM adjacent sequences that differ by 1, 2, 3 or more than 3 nucleotides from the selected target sites. Genomic DNA sequences for a target nucleic acid sequence, *e.g.*, a target gene may be obtained and repeat elements may be screened using publicly available tools, for example, the RepeatMasker program. RepeatMasker searches input DNA sequences for repeated elements and regions of low complexity. The output is a detailed annotation of the repeats present in a given query sequence.

Following identification, first regions of gRNAs, *e.g.*, crRNAs, are ranked into tiers based on their distance to the target site, their orthogonality and presence of 5' nucleotides for close matches with relevant PAM sequences (for example, a 5′ G based on identification of close matches in the human genome containing a relevant PAM *e.g.*, NGG PAM for S. pyogenes, NNGRRT or NNGRRV PAM for S. aureus). As used herein, orthogonality refers to the number of sequences in the human genome that contain a minimum number of

mismatches to the target sequence. A "high level of orthogonality" or "good orthogonality" may, for example, refer to 20-mer targeting domains that have no identical sequences in the human genome besides the intended target, nor any sequences that contain one or two mismatches in the target sequence. Targeting domains with good orthogonality may be selected to minimize off-target DNA cleavage.

A gRNA can then be introduced into a cell or embryo as an RNA molecule or a non-RNA nucleic acid molecule, *e.g.*, DNA molecule. In one embodiment, a DNA encoding a gRNA is operably linked to promoter control sequence for expression of the gRNA in a cell or embryo of interest. A RNA coding sequence can be operably linked to a promoter sequence that is recognized by RNA polymerase III (Pol III). Plasmid vectors that can be used to express gRNA include, but are not limited to, px330 vectors and px333 vectors. In some cases, a plasmid vector (*e.g.*, px333 vector) can comprise at least two gRNA-encoding DNA sequences. Further, a vector can comprise additional expression control sequences (*e.g.*, enhancer sequences, Kozak sequences, polyadenylation sequences, transcriptional termination sequences, etc.), selectable marker sequences (*e.g.*, GFP or antibiotic resistance genes such as puromycin), origins of replication, and the like. A DNA molecule encoding a gRNA can also be linear. A DNA molecule encoding a gRNA or a guide polynucleotide can also be circular.

In some embodiments, a reporter system is used for detecting base-editing activity and testing candidate guide polynucleotides. In some embodiments, a reporter system comprises a reporter gene based assay where base editing activity leads to expression of the reporter gene. For example, a reporter system may include a reporter gene comprising a deactivated start codon, *e.g.*, a mutation on the template strand from 3'-TAC-5' to 3'-CAC-5'. Upon successful deamination of the target C, the corresponding mRNA will be transcribed as 5'-AUG-3' instead of 5'-GUG-3', enabling the translation of the reporter gene. Suitable reporter genes will be apparent to those of skill in the art. Non-limiting examples of reporter genes include gene encoding green fluorescence protein (GFP), red fluorescence protein (RFP), luciferase, secreted alkaline phosphatase (SEAP), or any other gene whose expression are detectable and apparent to those skilled in the art. The reporter system can be used to test many different gRNAs, *e.g.*, in order to determine which residue(s) with respect to the target DNA sequence the respective deaminase will target. sgRNAs that target non-template strand can also be tested in order to assess off-target effects of a specific base editing protein, *e.g.*, a Cas9 deaminase fusion protein. In some embodiments, such gRNAs can be designed such that the mutated start codon will not be base-paired with the gRNA. The guide

polynucleotides can comprise standard ribonucleotides, modified ribonucleotides (*e.g.*, pseudouridine), ribonucleotide isomers, and/or ribonucleotide analogs. In some embodiments, the guide polynucleotide can comprise at least one detectable label. The detectable label can be a fluorophore (*e.g.*, FAM, TMR, Cy3, Cy5, Texas Red, Oregon Green, Alexa Fluors, Halo tags, or suitable fluorescent dye), a detection tag (*e.g.*, biotin, digoxigenin, and the like), quantum dots, or gold particles.

In some embodiments, a base editor system may comprise multiple guide polynucleotides, *e.g.*, gRNAs. For example, the gRNAs may target to one or more target loci (*e.g.*, at least 1 gRNA, at least 2 gRNA, at least 5 gRNA, at least 10 gRNA, at least 20 gRNA, at least 30 g RNA, at least 50 gRNA) comprised in a base editor system. The multiple gRNA sequences can be tandemly arranged and are preferably separated by a direct repeat.

Modified Polynucleotides

To enhance expression, stability, and/or genomic/base editing efficiency, and/or reduce possible toxicity, the base editor-coding sequence (e.g., mRNA) and/or the guide polynucleotide (e.g., gRNA) can be modified to include one or more modified nucleotides and/or chemical modifications, *e.g.* using pseudo-uridine, 5-Methyl-cytosine, 2′-O-methyl-3′-phosphonoacetate, 2′-*O*-methyl thioPACE (MSP), 2′-*O*-methyl-PACE (MP), 2′-fluoro RNA (2′-F-RNA), =constrained ethyl (S-cEt), 2′-O-methyl (‘M’), 2′-O-methyl-3′-phosphorothioate (‘MS’), 2′-O-methyl-3′-thiophosphonoacetate (‘MSP’), 5-methoxyuridine, phosphorothioate, and N1-Methylpseudouridine. Chemically protected gRNAs can enhance stability and editing efficiency *in vivo* and *ex vivo*. Methods for using chemically modified mRNAs and guide RNAs are known in the art and described, for example, by Jiang et al., Chemical modifications of adenine base editor mRNA and guide RNA expand its application scope. Nat Commun 11, 1979 (2020). doi.org/10.1038/s41467-020-15892-8, Callum et al., *N*1-Methylpseudouridine substitution enhances the performance of synthetic mRNA switches in cells, *Nucleic Acids Research*, Volume 48, Issue 6, 06 April 2020, Page e35, and Andries et al., Journal of Controlled Release, Volume 217, 10 November 2015, Pages 337-344, each of which is incorporated herein by reference in its entirety.

In a particular embodiment, the chemical modifications are 2′-*O*-methyl (2′-OMe) modifications. The modified guide RNAs may improve saCas9 efficacy and also specificity. The effect of an individual modification varies based on the position and combination of chemical modifications used as well as the inter- and intramolecular interactions with other

modified nucleotides. By way of example, S-cEt has been used to improve oligonucleotide intramolecular folding.

In some embodiments, the guide polynucleotide comprises one or more modified nucleotides at the 5' end and/or the 3' end of the guide. In some embodiments, the guide polynucleotide comprises two, three, four or more modified nucleosides at the 5' end and/or the 3' end of the guide. In some embodiments, the guide polynucleotide comprises two, three, four or more modified nucleosides at the 5' end and/or the 3' end of the guide. In some embodiments, the guide polynucleotide comprises four modified nucleosides at the 5' end and four modified nucleosides at the 3' end of the guide. In some embodiments, the modified nucleoside comprises a 2' O-methyl or a phosphorothioate.

In some embodiments, the guide comprises at least about 50%-75% modified nucleotides. In some embodiments, the guide comprises at least about 85% or more modified nucleotides. In some embodiments, at least about 1-5 nucleotides at the 5' end of the gRNA are modified and at least about 1-5 nucleotides at the 3' end of the gRNA are modified. In some embodiments, at least about 3-5 contiguous nucleotides at each of the 5' and 3' termini of the gRNA are modified. In some embodiments, at least about 20% of the nucleotides present in a direct repeat or anti-direct repeat are modified. In some embodiments, at least about 50% of the nucleotides present in a direct repeat or anti-direct repeat are modified. In some embodiments, at least about 50-75% of the nucleotides present in a direct repeat or anti-direct repeat are modified. In some embodiments, at least about 100 of the nucleotides present in a direct repeat or anti-direct repeat are modified. In some embodiments, at least about 20% or more of the nucleotides present in a hairpin present in the gRNA scaffold are modified. In some embodiments, at least about 50% or more of the nucleotides present in a hairpin present in the gRNA scaffold are modified. In some embodiments, the guide comprises a variable length spacer. In some embodiments, the guide comprises a 20-40 nucleotide spacer. In some embodiments, the guide comprises a spacer comprising at least about 20-25 nucleotides or at least about 30-35 nucleotides. In some embodiments, the spacer comprises modified nucleotides. In some embodiments, the guide comprises two or more of the following:

at least about 1-5 nucleotides at the 5' end of the gRNA are modified and at least about 1-5 nucleotides at the 3' end of the gRNA are modified;

at least about 20% of the nucleotides present in a direct repeat or anti-direct repeat are modified;

at least about 50-75% of the nucleotides present in a direct repeat or anti-direct repeat are modified;

at least about 20% or more of the nucleotides present in a hairpin present in the gRNA scaffold are modified;

a variable length spacer; and

a spacer comprising modified nucleotides.

In embodiments, the gRNA contains numerous modified nucleotides and/or chemical modifications ("heavy mods"). Such heavy mods can increase base editing ~2 fold *in vivo* or *in vitro*. For such modifications, mN = 2′-OMe**;** Ns = phosphorothioate (PS), where "N" represents the any nucleotide, as would be understood by one having skill in the art. In some cases, a nucleotide (N) may contain two modifications, for example, both a 2′-OMe and a PS modification. For example, a nucleotide with a phosphorothioate and 2' OMe is denoted as "mNs;" when there are two modifications next to each other, the notation is "mNsmNs.

In some embodiments of the modified gRNA, the gRNA comprises one or more chemical modifications selected from the group consisting of 2′-*O*-methyl (2′-OMe), phosphorothioate (PS), 2′-*O*-methyl thioPACE (MSP), 2′-*O*-methyl-PACE (MP), 2′-*O*-methyl thioPACE (MSP), 2′-fluoro RNA (2′-F-RNA), and constrained ethyl (S-cEt). In embodiments, the gRNA comprises 2′-*O*-methyl or phosphorothioate modifications. In an embodiment, the gRNA comprises 2′-*O*-methyl and phosphorothioate modifications. In an embodiment, the modifications increase base editing by at least about 2 fold.

A guide polynucleotide can comprise one or more modifications to provide a nucleic acid with a new or enhanced feature. A guide polynucleotide can comprise a nucleic acid affinity tag. A guide polynucleotide can comprise synthetic nucleotide, synthetic nucleotide analog, nucleotide derivatives, and/or modified nucleotides.

In some cases, a gRNA or a guide polynucleotide can comprise modifications. A modification can be made at any location of a gRNA or a guide polynucleotide. More than one modification can be made to a single gRNA or a guide polynucleotide. A gRNA or a guide polynucleotide can undergo quality control after a modification. In some cases, quality control can include PAGE, HPLC, MS, or any combination thereof.

A modification of a gRNA or a guide polynucleotide can be a substitution, insertion, deletion, chemical modification, physical modification, stabilization, purification, or any combination thereof.

A gRNA or a guide polynucleotide can also be modified by 5' adenylate, 5' guanosine-triphosphate cap, 5' N7-Methylguanosine-triphosphate cap, 5' triphosphate cap, 3'

phosphate, 3' thiophosphate, 5' phosphate, 5' thiophosphate, Cis-Syn thymidine dimer, trimers, C12 spacer, C3 spacer, C6 spacer, dSpacer, PC spacer, rSpacer, Spacer 18, Spacer 9, 3'-3' modifications, 2'-*O*-methyl thioPACE (MSP), 2'-*O*-methyl-PACE (MP), and constrained ethyl (S-cEt), 5'-5' modifications, abasic, acridine, azobenzene, biotin, biotin BB, biotin TEG, cholesteryl TEG, desthiobiotin TEG, DNP TEG, DNP-X, DOTA, dT-Biotin, dual biotin, PC biotin, psoralen C2, psoralen C6, TINA, 3' DABCYL, black hole quencher 1, black hole quencher 2, DABCYL SE, dT-DABCYL, IRDye QC-1, QSY-21, QSY-35, QSY-7, QSY-9, carboxyl linker, thiol linkers, 2'-deoxyribonucleoside analog purine, 2'-deoxyribonucleoside analog pyrimidine, ribonucleoside analog, 2'-O-methyl ribonucleoside analog, sugar modified analogs, wobble/universal bases, fluorescent dye label, 2'-fluoro RNA, 2'-O-methyl RNA, methylphosphonate, phosphodiester DNA, phosphodiester RNA, phosphothioate DNA, phosphorothioate RNA, UNA, pseudouridine-5'-triphosphate, 5'-methylcytidine-5'-triphosphate, or any combination thereof.

In some cases, a modification is permanent. In other cases, a modification is transient. In some cases, multiple modifications are made to a gRNA or a guide polynucleotide. A gRNA or a guide polynucleotide modification can alter physiochemical properties of a nucleotide, such as their conformation, polarity, hydrophobicity, chemical reactivity, base-pairing interactions, or any combination thereof.

A guide polynucleotide can be transferred into a cell by transfecting the cell with an isolated gRNA or a plasmid DNA comprising a sequence coding for the guide RNA and a promoter. A gRNA or a guide polynucleotide can also be transferred into a cell in other way, such as using virus-mediated gene delivery. A gRNA or a guide polynucleotide can be isolated. For example, a gRNA can be transfected in the form of an isolated RNA into a cell or organism. A gRNA can be prepared by *in vitro* transcription using any *in vitro* transcription system known in the art. A gRNA can be transferred to a cell in the form of isolated RNA rather than in the form of plasmid comprising encoding sequence for a gRNA.

A modification can also be a phosphorothioate substitute. In some cases, a natural phosphodiester bond can be susceptible to rapid degradation by cellular nucleases and; a modification of internucleotide linkage using phosphorothioate (PS) bond substitutes can be more stable towards hydrolysis by cellular degradation. A modification can increase stability in a gRNA or a guide polynucleotide. A modification can also enhance biological activity. In some cases, a phosphorothioate enhanced RNA gRNA can inhibit RNase A, RNase T1, calf serum nucleases, or any combinations thereof. These properties can allow the use of PS-RNA gRNAs to be used in applications where exposure to nucleases is of high probability *in*

vivo or *in vitro*. For example, phosphorothioate (PS) bonds can be introduced between the last 3-5 nucleotides at the 5'- or 3'-end of a gRNA which can inhibit exonuclease degradation. In some cases, phosphorothioate bonds can be added throughout an entire gRNA to reduce attack by endonucleases.

In some embodiments, the guide RNA is designed to disrupt a splice site (*i.e.*, a splice acceptor (SA) or a splice donor (SD). In some embodiments, the guide RNA is designed such that the base editing results in a premature STOP codon.

Protospacer Adjacent Motif

The term "protospacer adjacent motif (PAM)" or PAM-like motif refers to a 2-6 base pair DNA sequence immediately following the DNA sequence targeted by the Cas9 nuclease in the CRISPR bacterial adaptive immune system. In some embodiments, the PAM can be a 5' PAM (*i.e.*, located upstream of the 5' end of the protospacer). In other embodiments, the PAM can be a 3' PAM (*i.e.*, located downstream of the 5' end of the protospacer). The PAM sequence is essential for target binding, but the exact sequence depends on a type of Cas protein. The PAM sequence can be any PAM sequence known in the art. Suitable PAM sequences include, but are not limited to, NGG, NGA, NGC, NGN, NGT, NGTT, NGCG, NGAG, NGAN, NGNG, NGCN, NGCG, NGTN, NNGRRT, NNNRRT, NNGRR(N), TTTV, TYCV, TYCV, TATV, NNNNGATT, NNAGAAW, or NAAAAC. Y is a pyrimidine; N is any nucleotide base; W is A or T.

A base editor provided herein can comprise a CRISPR protein-derived domain that is capable of binding a nucleotide sequence that contains a canonical or non-canonical protospacer adjacent motif (PAM) sequence. A PAM site is a nucleotide sequence in proximity to a target polynucleotide sequence. Some aspects of the disclosure provide for base editors comprising all or a portion of CRISPR proteins that have different PAM specificities.

For example, typically Cas9 proteins, such as Cas9 from *S. pyogenes* (spCas9), require a canonical NGG PAM sequence to bind a particular nucleic acid region, where the "N" in "NGG" is adenine (A), thymine (T), guanine (G), or cytosine (C), and the G is guanine. A PAM can be CRISPR protein-specific and can be different between different base editors comprising different CRISPR protein-derived domains. A PAM can be 5' or 3' of a target sequence. A PAM can be upstream or downstream of a target sequence. A PAM can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more nucleotides in length. Often, a PAM is between 2-6 nucleotides in length.

In some embodiments, the PAM is an "NRN" PAM where the "N" in "NRN" is adenine (A), thymine (T), guanine (G), or cytosine (C), and the R is adenine (A) or guanine (G); or the PAM is an "NYN" PAM, wherein the "N" in NYN is adenine (A), thymine (T), guanine (G), or cytosine (C), and the Y is cytidine (C) or thymine (T), for example, as described in R.T. Walton *et al.*, 2020, *Science*, 10.1126/science.aba8853 (2020), the entire contents of which are incorporated herein by reference.

Several PAM variants are described in **Table 6** below.

Table 6. Cas9 proteins and corresponding PAM sequences

Variant	PAM
spCas9	NGG
spCas9-VRQR	NGA
spCas9-VRER	NGCG
xCas9 (sp)	NGN
saCas9	NNGRRT
saCas9-KKH	NNNRRT
spCas9-MQKSER	NGCG
spCas9-MQKSER	NGCN
spCas9-LRKIQK	NGTN
spCas9-LRVSQK	NGTN
spCas9-LRVSQL	NGTN
spCas9-MQKFRAER	NGC
Cpf1	5' (TTTV)
SpyMac	5'-NAA-3'

In some embodiments, the PAM is NGC. In some embodiments, the NGC PAM is recognized by a Cas9 variant. In some embodiments, the NGC PAM variant includes one or more amino acid substitutions selected from D1135M, S1136Q, G1218K, E1219F, A1322R, D1332A, R1335E, and T1337R (collectively termed "MQKFRAER").

In some embodiments, the PAM is NGT. In some embodiments, the NGT PAM is recognized by a Cas9 variant. In some embodiments, the NGT PAM variant is generated through targeted mutations at one or more residues 1335, 1337, 1135, 1136, 1218, and/or

1219. In some embodiments, the NGT PAM variant is created through targeted mutations at one or more residues 1219, 1335, 1337, 1218. In some embodiments, the NGT PAM variant is created through targeted mutations at one or more residues 1135, 1136, 1218, 1219, and 1335. In some embodiments, the NGT PAM variant is selected from the set of targeted mutations provided in **Tables 7A** and **7B** below.

Table 7A: NGT PAM Variant Mutations at residues 1219, 1335, 1337, 1218

Variant	E1219V	R1335Q	T1337	G1218
1	F	V	T	
2	F	V	R	
3	F	V	Q	
4	F	V	L	
5	F	V	T	R
6	F	V	R	R
7	F	V	Q	R
8	F	V	L	R
9	L	L	T	
10	L	L	R	
11	L	L	Q	
12	L	L	L	
13	F	I	T	
14	F	I	R	
15	F	I	Q	
16	F	I	L	
17	F	G	C	
18	H	L	N	
19	F	G	C	A
20	H	L	N	V
21	L	A	W	
22	L	A	F	
23	L	A	Y	
24	I	A	W	
25	I	A	F	
26	I	A	Y	

Table 7B: NGT PAM Variant Mutations at residues 1135, 1136, 1218, 1219, and 1335

Variant	D1135L	S1136R	G1218S	E1219V	R1335Q
27	G				
28	V				
29	I				
30		A			
31		W			
32		H			
33		K			

34			K		
35			R		
36			Q		
37			T		
38			N		
39				I	
40				A	
41				N	
42				Q	
43				G	
44				L	
45				S	
46				T	
47					L
48					I
49					V
50					N
51					S
52					T
53					F
54					Y
55	N1286Q	I1331F			

In some embodiments, the NGT PAM variant is selected from variant 5, 7, 28, 31, or 36 in **Table 7A** and **Table 7B**. In some embodiments, the variants have improved NGT PAM recognition.

In some embodiments, the NGT PAM variants have mutations at residues 1219, 1335, 1337, and/or 1218. In some embodiments, the NGT PAM variant is selected with mutations for improved recognition from the variants provided in **Table 8** below.

Table 8: NGT PAM Variant Mutations at residues 1219, 1335, 1337, and 1218

Variant	E1219V	R1335Q	T1337	G1218
1	F	V	T	
2	F	V	R	
3	F	V	Q	
4	F	V	L	
5	F	V	T	R
6	F	V	R	R
7	F	V	Q	R
8	F	V	L	R

In some embodiments, the NGT PAM is selected from the variants provided in **Table 9** below.

Table 9. NGT PAM variants

	NGTN variant	D1135	S1136	G1218	E1219	A1322R	R1335	T1337
Variant 1	LRKIQK	L	R	K	I	-	Q	K
Variant 2	LRSVQK	L	R	S	V	-	Q	K
Variant 3	LRSVQL	L	R	S	V	-	Q	L
Variant 4	LRKIRQK	L	R	K	I	R	Q	K
Variant 5	LRSVRQK	L	R	S	V	R	Q	K
Variant 6	LRSVRQL	L	R	S	V	R	Q	L

In some embodiments the NGTN variant is variant 1. In some embodiments, the NGTN variant is variant 2. In some embodiments, the NGTN variant is variant 3. In some embodiments, the NGTN variant is variant 4. In some embodiments, the NGTN variant is variant 5. In some embodiments, the NGTN variant is variant 6.

In some embodiments, the Cas9 domain is a Cas9 domain from *Streptococcus pyogenes* (SpCas9). In some embodiments, the SpCas9 domain is a nuclease active SpCas9, a nuclease inactive SpCas9 (SpCas9d), or a SpCas9 nickase (SpCas9n). In some embodiments, the SpCas9 comprises a D9X mutation, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid except for D. In some embodiments, the SpCas9 comprises a D9A mutation, or a corresponding mutation in any of the amino acid sequences provided herein. In some embodiments, the SpCas9 domain, the SpCas9d domain, or the SpCas9n domain can bind to a nucleic acid sequence having a non-canonical PAM. In some embodiments, the SpCas9 domain, the SpCas9d domain, or the SpCas9n domain can bind to a nucleic acid sequence having an NGG, a NGA, or a NGCG PAM sequence.

In some embodiments, the SpCas9 domain comprises one or more of a D1135X, a R1335X, and a T1337X mutation, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid. In some embodiments, the SpCas9 domain comprises one or more of a D1135E, R1335Q, and T1337R mutation, or a corresponding mutation in any of the amino acid sequences provided herein. In some embodiments, the SpCas9 domain comprises a D1135E, a R1335Q, and a T1337R mutation, or corresponding mutations in any of the amino acid sequences provided herein. In some embodiments, the SpCas9 domain comprises one or more of a D1135X, a R1335X, and a T1337X mutation, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid. In some embodiments, the SpCas9 domain comprises one or more of a D1135V, a R1335Q, and a T1337R mutation, or a corresponding mutation

in any of the amino acid sequences provided herein. In some embodiments, the SpCas9 domain comprises a D1135V, a R1335Q, and a T1337R mutation, or corresponding mutations in any of the amino acid sequences provided herein. In some embodiments, the SpCas9 domain comprises one or more of a D1135X, a G1218X, a R1335X, and a T1337X mutation, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid. In some embodiments, the SpCas9 domain comprises one or more of a D1135V, a G1218R, a R1335Q, and a T1337R mutation, or a corresponding mutation in any of the amino acid sequences provided herein. In some embodiments, the SpCas9 domain comprises a D1135V, a G1218R, a R1335Q, and a T1337R mutation, or corresponding mutations in any of the amino acid sequences provided herein.

In some examples, a PAM recognized by a CRISPR protein-derived domain of a base editor disclosed herein can be provided to a cell on a separate oligonucleotide to an insert (*e.g.*, an AAV insert) encoding the base editor. In such embodiments, providing PAM on a separate oligonucleotide can allow cleavage of a target sequence that otherwise would not be able to be cleaved, because no adjacent PAM is present on the same polynucleotide as the target sequence.

In an embodiment, *S. pyogenes* Cas9 (SpCas9) can be used as a CRISPR endonuclease for genome engineering. However, others can be used. In some embodiments, a different endonuclease can be used to target certain genomic targets. In some embodiments, synthetic SpCas9-derived variants with non-NGG PAM sequences can be used. Additionally, other Cas9 orthologues from various species have been identified and these "non-SpCas9s" can bind a variety of PAM sequences that can also be useful for the present disclosure. For example, the relatively large size of SpCas9 (approximately 4kb coding sequence) can lead to plasmids carrying the SpCas9 cDNA that cannot be efficiently expressed in a cell. Conversely, the coding sequence for *Staphylococcus aureus* Cas9 (SaCas9) is approximately 1 kilobase shorter than SpCas9, possibly allowing it to be efficiently expressed in a cell. Similar to SpCas9, the SaCas9 endonuclease is capable of modifying target genes in mammalian cells *in vitro* and in mice *in vivo*. In some embodiments, a Cas protein can target a different PAM sequence. In some embodiments, a target gene can be adjacent to a Cas9 PAM, 5'-NGG, for example. In other embodiments, other Cas9 orthologs can have different PAM requirements. For example, other PAMs such as those of *S. thermophilus* (5'-NNAGAA for CRISPR1 and 5'-NGGNG for CRISPR3) and *Neisseria meningitidis* (5'-NNNNGATT) can also be found adjacent to a target gene.

In some embodiments, for a *S. pyogenes* system, a target gene sequence can precede (*i.e.*, be 5' to) a 5'-NGG PAM, and a 20-nt guide RNA sequence can base pair with an opposite strand to mediate a Cas9 cleavage adjacent to a PAM. In some embodiments, an adjacent cut can be or can be about 3 base pairs upstream of a PAM. In some embodiments, an adjacent cut can be or can be about 10 base pairs upstream of a PAM. In some embodiments, an adjacent cut can be or can be about 0-20 base pairs upstream of a PAM. For example, an adjacent cut can be next to, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 base pairs upstream of a PAM. An adjacent cut can also be downstream of a PAM by 1 to 30 base pairs. The sequences of exemplary SpCas9 proteins capable of binding a PAM sequence follow:

In some embodiments, engineered SpCas9 variants are capable of recognizing protospacer adjacent motif (PAM) sequences flanked by a 3′ H (non-G PAM) (*see* Tables 2A-2D). In some embodiments, the SpCas9 variants recognize NRNH PAMs (where R is A or G and H is A, C or T). In some embodiments, the non-G PAM is NRRH, NRTH, or NRCH (*see e.g.*, Miller, S.M., *et al.* Continuous evolution of SpCas9 variants compatible with non-G PAMs, Nat. Biotechnol. (2020), the contents of which is incorporated herein by reference in its entirety).

In some embodiments, the Cas9 domain is a recombinant Cas9 domain. In some embodiments, the recombinant Cas9 domain is a SpyMacCas9 domain. In some embodiments, the SpyMacCas9 domain is a nuclease active SpyMacCas9, a nuclease inactive SpyMacCas9 (SpyMacCas9d), or a SpyMacCas9 nickase (SpyMacCas9n). In some embodiments, the SaCas9 domain, the SaCas9d domain, or the SaCas9n domain can bind to a nucleic acid sequence having a non-canonical PAM. In some embodiments, the SpyMacCas9 domain, the SpCas9d domain, or the SpCas9n domain can bind to a nucleic acid sequence having a NAA PAM sequence.

The sequence of an exemplary Cas9 A homolog of Spy Cas9 in *Streptococcus macacae* with native 5'-NAAN-3' PAM specificity is known in the art and described, for example, by Chatterjee, *et al.*, "A Cas9 with PAM recognition for adenine dinucleotides", *Nature Communications*, vol. 11, article no. 2474 (2020), and is in the Sequence Listing as SEQ ID NO: 325.

In some embodiments, a variant Cas9 protein harbors, H840A, P475A, W476A, N477A, D1125A, W1126A, and D1218A mutations such that the polypeptide has a reduced ability to cleave a target DNA or RNA. Such a Cas9 protein has a reduced ability to cleave a

target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a target DNA (*e.g.*, a single stranded target DNA). As another non-limiting example, in some embodiments, the variant Cas9 protein harbors D10A, H840A, P475A, W476A, N477A, D1125A, W1126A, and D1218A mutations such that the polypeptide has a reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a target DNA (*e.g.*, a single stranded target DNA). In some embodiments, when a variant Cas9 protein harbors W476A and W1126A mutations or when the variant Cas9 protein harbors P475A, W476A, N477A, D1125A, W1126A, and D1218A mutations, the variant Cas9 protein does not bind efficiently to a PAM sequence. Thus, in some such cases, when such a variant Cas9 protein is used in a method of binding, the method does not require a PAM sequence. In other words, in some embodiments, when such a variant Cas9 protein is used in a method of binding, the method can include a guide RNA, but the method can be performed in the absence of a PAM sequence (and the specificity of binding is therefore provided by the targeting segment of the guide RNA). Other residues can be mutated to achieve the above effects (*i.e.*, inactivate one or the other nuclease portions). As non-limiting examples, residues D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, and/or A987 can be altered (*i.e.*, substituted). Also, mutations other than alanine substitutions are suitable.

In some embodiments, a CRISPR protein-derived domain of a base editor can comprise all or a portion of a Cas9 protein with a canonical PAM sequence (NGG). In other embodiments, a Cas9-derived domain of a base editor can employ a non-canonical PAM sequence. Such sequences have been described in the art and would be apparent to the skilled artisan. For example, Cas9 domains that bind non-canonical PAM sequences have been described in Kleinstiver, B. P., *et al.*, "Engineered CRISPR-Cas9 nucleases with altered PAM specificities" Nature 523, 481-485 (2015); and Kleinstiver, B. P., *et al.*, "Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition" Nature Biotechnology 33, 1293-1298 (2015); R.T. Walton *et al.* "Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants" *Science* 10.1126/science.aba8853 (2020); Hu *et al.* "Evolved Cas9 variants with broad PAM compatibility and high DNA specificity," *Nature*, 2018 Apr. 5, 556(7699), 57-63; Miller *et al.*, "Continuous evolution of SpCas9 variants compatible with non-G PAMs" *Nat. Biotechnol.*, 2020 Apr;38(4):471-481; the entire contents of each are hereby incorporated by reference.

Fusion Proteins Comprising a NapDNAbp and a Cytidine Deaminase and/or Adenosine Deaminase

Some aspects of the disclosure provide fusion proteins comprising a Cas9 domain or other nucleic acid programmable DNA binding protein (*e.g.*, Cas12) and one or more cytidine deaminase or adenosine deaminase domains. It should be appreciated that the Cas9 domain may be any of the Cas9 domains or Cas9 proteins (*e.g.*, dCas9 or nCas9) provided herein. In some embodiments, any of the Cas9 domains or Cas9 proteins (*e.g.*, dCas9 or nCas9) provided herein may be fused with any of the cytidine deaminases and/or adenosine deaminases provided herein. The domains of the base editors disclosed herein can be arranged in any order.

In some embodiments, the fusion protein comprises the following domains A-C, A-D, or A-E:

NH_2-[A-B-C]-COOH;

NH_2-[A-B-C-D]-COOH; or

NH_2-[A-B-C-D-E]-COOH;

wherein A and C or A, C, and E, each comprises one or more of the following:

an adenosine deaminase domain or an active fragment thereof,

a cytidine deaminase domain or an active fragment thereof; and

wherein B or B and D, each comprises one or more domains having nucleic acid sequence specific binding activity.

In some embodiments, the fusion protein comprises the following structure:

NH_2-[A_n-B_o-C_n]-COOH;

NH_2-[A_n-B_o-C_n-D_o]-COOH; or

NH_2-[A_n-B_o-C_p-D_o-E_q]-COOH;

wherein A and C or A, C, and E, each comprises one or more of the following:

an adenosine deaminase domain or an active fragment thereof,

a cytidine deaminase domain or an active fragment thereof; and

wherein n is an integer: 1, 2, 3, 4, or 5, wherein p is an integer: 0, 1, 2, 3, 4, or 5; wherein q is an integer 0, 1, 2, 3, 4, or 5; and wherein B or B and D each comprises a domain having nucleic acid sequence specific binding activity; and wherein o is an integer: 1, 2, 3, 4, or 5.

For example, and without limitation, in some embodiments, the fusion protein comprises the structure:

NH2-[adenosine deaminase]-[Cas9 domain]-COOH;

NH2-[Cas9 domain]-[adenosine deaminase]-COOH;

NH2-[cytidine deaminase]-[Cas9 domain]-COOH;
NH2-[Cas9 domain]-[cytidine deaminase]-COOH;
NH2-[cytidine deaminase]-[Cas9 domain]-[adenosine deaminase]-COOH;
NH2-[adenosine deaminase]-[Cas9 domain]-[cytidine deaminase]-COOH;
NH2-[adenosine deaminase]-[cytidine deaminase]-[Cas9 domain]-COOH;
NH2-[cytidine deaminase]-[adenosine deaminase]-[Cas9 domain]-COOH;
NH2-[Cas9 domain]-[adenosine deaminase]-[cytidine deaminase]-COOH; or
NH2-[Cas9 domain]-[cytidine deaminase]-[adenosine deaminase]-COOH.

In some embodiments, any of the Cas12 domains or Cas12 proteins provided herein may be fused with any of the cytidine or adenosine deaminases provided herein. For example, and without limitation, in some embodiments, the fusion protein comprises the structure:

NH2-[adenosine deaminase]-[Cas12 domain]-COOH;
NH2-[Cas12 domain]-[adenosine deaminase]-COOH;
NH2-[cytidine deaminase]-[Cas12 domain]-COOH;
NH2-[Cas12 domain]-[cytidine deaminase]-COOH;
NH2-[cytidine deaminase]-[Cas12 domain]-[adenosine deaminase]-COOH;
NH2-[adenosine deaminase]-[Cas12 domain]-[cytidine deaminase]-COOH;
NH2-[adenosine deaminase]-[cytidine deaminase]-[Cas12 domain]-COOH;
NH2-[cytidine deaminase]-[adenosine deaminase]-[Cas12 domain]-COOH;
NH2-[Cas12 domain]-[adenosine deaminase]-[cytidine deaminase]-COOH; or
NH2-[Cas12 domain]-[cytidine deaminase]-[adenosine deaminase]-COOH.

In some embodiments, the adenosine deaminase is a TadA*8. Exemplary fusion protein structures include the following:

NH2-[TadA*8]-[Cas9 domain]-COOH;
NH2-[Cas9 domain]-[TadA*8]-COOH;
NH2-[TadA*8]-[Cas12 domain]-COOH; or
NH2-[Cas12 domain]-[TadA*8]-COOH.

In some embodiments, the adenosine deaminase of the fusion protein comprises a TadA*8 and a cytidine deaminase and/or an adenosine deaminase. In some embodiments, the TadA*8 is TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, TadA*8.7, TadA*8.8, TadA*8.9, TadA*8.10, TadA*8.11, TadA*8.12, TadA*8.13, TadA*8.14, TadA*8.15, TadA*8.16, TadA*8.17, TadA*8.18, TadA*8.19, TadA*8.20, TadA*8.21, TadA*8.22, TadA*8.23, or TadA*8.24.

Exemplary fusion protein structures include the following:

NH2-[TadA*8]-[Cas9/Cas12]-[adenosine deaminase]-COOH;

NH2-[adenosine deaminase]-[Cas9/Cas12]-[TadA*8]-COOH;

NH2-[TadA*8]-[Cas9/Cas12]-[cytidine deaminase]-COOH; or

NH2-[cytidine deaminase]-[Cas9/Cas12]-[TadA*8]-COOH.

In some embodiments, the adenosine deaminase of the fusion protein comprises a TadA*9 and a cytidine deaminase and/or an adenosine deaminase. Exemplary fusion protein structures include the following:

NH2-[TadA*9]-[Cas9/Cas12]-[adenosine deaminase]-COOH;

NH2-[adenosine deaminase]-[Cas9/Cas12]-[TadA*9]-COOH;

NH2-[TadA*9]-[Cas9/Cas12]-[cytidine deaminase]-COOH; or

NH2-[cytidine deaminase]-[Cas9/Cas12]-[TadA*9]-COOH.

In some embodiments, the fusion protein can comprise a deaminase flanked by an N-terminal fragment and a C-terminal fragment of a Cas9 or Cas12 polypeptide. In some embodiments, the fusion protein comprises a cytidine deaminase flanked by an N- terminal fragment and a C-terminal fragment of a Cas9 or Cas12 polypeptide. In some embodiments, the fusion protein comprises an adenosine deaminase flanked by an N- terminal fragment and a C-terminal fragment of a Cas9 or Cas 12 polypeptide.

In some embodiments, the fusion proteins comprising a cytidine deaminase or adenosine deaminase and a napDNAbp (*e.g.*, Cas9 or Cas12 domain) do not include a linker sequence. In some embodiments, a linker is present between the cytidine or adenosine deaminase and the napDNAbp. In some embodiments, the "-" used in the general architecture above indicates the presence of an optional linker. In some embodiments, cytidine or adenosine deaminase and the napDNAbp are fused via any of the linkers provided herein. For example, in some embodiments the cytidine or adenosine deaminase and the napDNAbp are fused via any of the linkers provided herein.

It should be appreciated that the fusion proteins of the present disclosure may comprise one or more additional features. For example, in some embodiments, the fusion protein may comprise inhibitors, cytoplasmic localization sequences, export sequences, such as nuclear export sequences, or other localization sequences, as well as sequence tags that are useful for solubilization, purification, or detection of the fusion proteins. Suitable protein tags provided herein include, but are not limited to, biotin carboxylase carrier protein (BCCP) tags, myc-tags, calmodulin-tags, FLAG-tags, hemagglutinin (HA)-tags, polyhistidine tags, also referred to as histidine tags or His-tags, maltose binding protein (MBP)-tags, nus-tags,

glutathione-S-transferase (GST)-tags, green fluorescent protein (GFP)-tags, thioredoxin-tags, S-tags, Softags (*e.g.*, Softag 1, Softag 3), strep-tags , biotin ligase tags, FlAsH tags, V5 tags, and SBP-tags. Additional suitable sequences will be apparent to those of skill in the art. In some embodiments, the fusion protein comprises one or more His tags.

Exemplary, yet nonlimiting, fusion proteins are described in International PCT Application Nos. PCT/2017/044935, PCT/US2019/044935, and PCT/US2020/016288, each of which is incorporated herein by reference for its entirety.

Fusion Proteins Comprising a Nuclear Localization Sequence (NLS)

In some embodiments, the fusion proteins provided herein further comprise one or more (*e.g.*, 2, 3, 4, 5) nuclear targeting sequences, for example a nuclear localization sequence (NLS). In one embodiment, a bipartite NLS is used. In some embodiments, a NLS comprises an amino acid sequence that facilitates the importation of a protein, that comprises an NLS, into the cell nucleus (*e.g.*, by nuclear transport). In some embodiments, the NLS is fused to the N-terminus or the C-terminus of the fusion protein. In some embodiments, the NLS is fused to the C-terminus or N-terminus of an nCas9 domain or a dCas9 domain. In some embodiments, the NLS is fused to the N-terminus or C-terminus of the Cas12 domain. In some embodiments, the NLS is fused to the N-terminus or C-terminus of the cytidine or adenosine deaminase. In some embodiments, the NLS is fused to the fusion protein via one or more linkers. In some embodiments, the NLS is fused to the fusion protein without a linker. In some embodiments, the NLS comprises an amino acid sequence of any one of the NLS sequences provided or referenced herein. Additional nuclear localization sequences are known in the art and would be apparent to the skilled artisan. For example, NLS sequences are described in Plank *et al.*, PCT/EP2000/011690, the contents of which are incorporated herein by reference for their disclosure of exemplary nuclear localization sequences. In some embodiments, an NLS comprises the amino acid sequence PKKKRKVEGADKRTADGSEFESPKKKRKV (SEQ ID NO: 416), KRTADGSEFESPKKKRKV (SEQ ID NO: 243), KRPAATKKAGQAKKKK (SEQ ID NO: 244), KKTELQTTNAENKTKKL (SEQ ID NO: 245), KRGINDRNFWRGENGRKTR (SEQ ID NO: 246), RKSGKIAAIVVKRPRKPKKKRKV (SEQ ID NO: 417), or MDSLLMNRRKFLYQFKNVRWAKGRRETYLC (SEQ ID NO: 249).

In some embodiments, the fusion proteins comprising a cytidine or adenosine deaminase, a Cas9 domain, and an NLS do not comprise a linker sequence. In some embodiments, linker sequences between one or more of the domains or proteins (*e.g.*,

cytidine or adenosine deaminase, Cas9 domain or NLS) are present. In some embodiments, a linker is present between the cytidine deaminase and adenosine deaminase domains and the napDNAbp. In some embodiments, the "-" used in the general architecture below indicates the presence of an optional linker. In some embodiments, the cytidine deaminase and adenosine deaminase and the napDNAbp are fused via any of the linkers provided herein. For example, in some embodiments the cytidine deaminase and adenosine deaminase and the napDNAbp are fused via any of the linkers provided herein.

In some embodiments, the general architecture of exemplary napDNAbp (*e.g.*, Cas9 or Cas12) fusion proteins with a cytidine or adenosine deaminase and a napDNAbp (*e.g.*, Cas9 or Cas12) domain comprises any one of the following structures, where NLS is a nuclear localization sequence (*e.g.*, any NLS provided herein), NH_2 is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein:

NH_2-NLS-[cytidine deaminase]-[napDNAbp domain]-COOH;

NH_2-NLS [napDNAbp domain]-[cytidine deaminase]-COOH;

NH_2-[cytidine deaminase]-[napDNAbp domain]-NLS-COOH;

NH_2-[napDNAbp domain]-[cytidine deaminase]-NLS-COOH;

NH_2-NLS-[adenosine deaminase]-[napDNAbp domain]-COOH;

NH_2-NLS [napDNAbp domain]-[adenosine deaminase]-COOH;

NH_2-[adenosine deaminase]-[napDNAbp domain]-NLS-COOH;

NH_2-[napDNAbp domain]-[adenosine deaminase]-NLS-COOH;

NH_2-NLS-[cytidine deaminase]-[napDNAbp domain]-[adenosine deaminase]-COOH;

NH_2-NLS-[adenosine deaminase]-[napDNAbp domain]-[cytidine deaminase]-COOH;

NH_2-NLS-[adenosine deaminase] [cytidine deaminase]-[napDNAbp domain]-COOH;

NH_2-NLS-[cytidine deaminase]-[adenosine deaminase]-[napDNAbp domain]-COOH;

NH_2-NLS-[napDNAbp domain]-[adenosine deaminase]-[cytidine deaminase]-COOH;

NH_2-NLS-[napDNAbp domain]-[cytidine deaminase]-[adenosine deaminase]-COOH;

NH_2-[cytidine deaminase]-[napDNAbp domain]-[adenosine deaminase]-NLS-COOH;

NH_2-[adenosine deaminase]-[napDNAbp domain]-[cytidine deaminase]-NLS-COOH;

NH_2-[adenosine deaminase] [cytidine deaminase]-[napDNAbp domain]-NLS-COOH;

NH_2-[cytidine deaminase]-[adenosine deaminase]-[napDNAbp domain]-NLS-COOH;

NH_2-[napDNAbp domain]-[adenosine deaminase]-[cytidine deaminase]-NLS-COOH; or

NH_2-[napDNAbp domain]-[cytidine deaminase]-[adenosine deaminase]-NLS-COOH.

In some embodiments, the NLS is present in a linker or the NLS is flanked by linkers, for example described herein. A bipartite NLS comprises two basic amino acid clusters, which

are separated by a relatively short spacer sequence (hence bipartite - 2 parts, while monopartite NLSs are not). The NLS of nucleoplasmin, KR[PAATKKAGQA]KKKK (SEQ ID NO: 244), is the prototype of the ubiquitous bipartite signal: two clusters of basic amino acids, separated by a spacer of about 10 amino acids. The sequence of an exemplary bipartite NLS follows: PKKKRKVEGADKRTADGSEFESPKKKRKV (SEQ ID NO: 416).

A vector that encodes a CRISPR enzyme comprising one or more nuclear localization sequences (NLSs) can be used. For example, there can be or be about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 NLSs used. A CRISPR enzyme can comprise the NLSs at or near the amino-terminus, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 NLSs at or near the carboxy-terminus, or any combination thereof (*e.g.*, one or more NLS at the amino-terminus and one or more NLS at the carboxy terminus). When more than one NLS is present, each can be selected independently of others, such that a single NLS can be present in more than one copy and/or in combination with one or more other NLSs present in one or more copies.

CRISPR enzymes used in the methods can comprise about 6 NLSs. An NLS is considered near the N- or C-terminus when the nearest amino acid to the NLS is within about 50 amino acids along a polypeptide chain from the N- or C-terminus, *e.g.*, within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, or 50 amino acids.

Additional Domains

A base editor described herein can include any domain which helps to facilitate the nucleobase editing, modification or altering of a nucleobase of a polynucleotide. In various embodiments, open reading frames encoding any of these additional domains may be modified to include an intron subject to inactivation according to the methods described herein. In some embodiments, a base editor comprises a polynucleotide programmable nucleotide binding domain (*e.g.*, Cas9), a nucleobase editing domain (*e.g.*, deaminase domain), and one or more additional domains. In some embodiments, the additional domain can facilitate enzymatic or catalytic functions of the base editor, binding functions of the base editor, or be inhibitors of cellular machinery (*e.g.*, enzymes) that could interfere with the desired base editing result. In some embodiments, a base editor can comprise a nuclease, a nickase, a recombinase, a deaminase, a methyltransferase, a methylase, an acetylase, an acetyltransferase, a transcriptional activator, or a transcriptional repressor domain.

In some embodiments, a base editor can comprise an uracil glycosylase inhibitor (UGI) domain. In some embodiments, cellular DNA repair response to the presence of U: G heteroduplex DNA can be responsible for a decrease in nucleobase editing efficiency in cells.

In such embodiments, uracil DNA glycosylase (UDG) can catalyze removal of U from DNA in cells, which can initiate base excision repair (BER), mostly resulting in reversion of the U:G pair to a C:G pair. In such embodiments, BER can be inhibited in base editors comprising one or more domains that bind the single strand, block the edited base, inhibit UGI, inhibit BER, protect the edited base, and /or promote repairing of the non-edited strand. Thus, this disclosure contemplates a base editor fusion protein comprising a UGI domain.

In some embodiments, a base editor comprises as a domain all or a portion of a double-strand break (DSB) binding protein. For example, a DSB binding protein can include a Gam protein of bacteriophage Mu that can bind to the ends of DSBs and can protect them from degradation. See Komor, A.C., *et al.*, "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017), the entire content of which is hereby incorporated by reference.

Additionally, in some embodiments, a Gam protein can be fused to an N terminus of a base editor. In some embodiments, a Gam protein can be fused to a C terminus of a base editor. The Gam protein of bacteriophage Mu can bind to the ends of double strand breaks (DSBs) and protect them from degradation. In some embodiments, using Gam to bind the free ends of DSB can reduce indel formation during the process of base editing. In some embodiments, 174-residue Gam protein is fused to the N terminus of the base editors. See Komor, A.C., *et al.*, "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017). In some embodiments, a mutation or mutations can change the length of a base editor domain relative to a wild type domain. For example, a deletion of at least one amino acid in at least one domain can reduce the length of the base editor. In another case, a mutation or mutations do not change the length of a domain relative to a wild type domain. For example, substitutions in any domain does not change the length of the base editor.

Non-limiting examples of such base editors, where the length of all the domains is the same as the wild type domains, can include:

NH2-[nucleobase editing domain]-Linker1-[APOBEC1]-Linker2-[nucleobase editing domain]-COOH;

NH2-[nucleobase editing domain]-Linker1-[APOBEC1]-[nucleobase editing domain]-COOH;

NH2-[nucleobase editing domain]-[APOBEC1]-Linker2-[nucleobase editing domain]-COOH;

NH2-[nucleobase editing domain]-[APOBEC1]-[nucleobase editing domain]-COOH;

NH2-[nucleobase editing domain]-Linker1-[APOBEC1]-Linker2-[nucleobase editing domain]-[UGI]-COOH;

NH2-[nucleobase editing domain]-Linker1-[APOBEC1]-[nucleobase editing domain]-[UGI]-COOH;

NH2-[nucleobase editing domain]-[APOBEC1]-Linker2-[nucleobase editing domain]-[UGI]-COOH;

NH2-[nucleobase editing domain]-[APOBEC1]-[nucleobase editing domain]-[UGI]-COOH;

NH2-[UGI]-[nucleobase editing domain]-Linker1-[APOBEC1]-Linker2-[nucleobase editing domain]-COOH;

NH2-[UGI]-[nucleobase editing domain]-Linker1-[APOBEC1]-[nucleobase editing domain]-COOH;

NH2-[UGI]-[nucleobase editing domain]-[APOBEC1]-Linker2-[nucleobase editing domain]-COOH; or

NH2-[UGI]-[nucleobase editing domain]-[APOBEC1]-[nucleobase editing domain]-COOH.

BASE EDITOR SYSTEM

Provided herein are systems, compositions, and methods for editing a nucleobase using a base editor system featuring a self-inactivating base editor. In some embodiments, the base editor system comprises (1) a base editor (BE) comprising a polynucleotide programmable nucleotide binding domain and a nucleobase editing domain (*e.g.*, a deaminase domain) for editing the nucleobase; and (2) a guide polynucleotide (*e.g.*, guide RNA) in conjunction with the polynucleotide programmable nucleotide binding domain. In some embodiments, the base editor system is a cytidine base editor (CBE) or an adenosine base editor (ABE). Introns can be inserted into an open reading frame encoding a polynucleotide programmable nucleotide binding domain, a nucleobase editing domain, or a fragment of one of those domains. In some embodiments, the polynucleotide programmable nucleotide binding domain is a polynucleotide programmable DNA or RNA binding domain. In some embodiments, the nucleobase editing domain is a deaminase domain. In some embodiments, a deaminase domain can be a cytidine deaminase or an cytosine deaminase. In some embodiments, a deaminase domain can be an adenine deaminase or an adenosine deaminase.

In some embodiments, the adenosine base editor can deaminate adenine in DNA. In some embodiments, the base editor is capable of deaminating a cytidine in DNA.

In some embodiments, a base editing system as provided herein provides a new approach to genome editing that uses a fusion protein containing a catalytically defective *Streptococcus pyogenes* Cas9, a deaminase (*e.g.*, cytidine or adenosine deaminase), and an inhibitor of base excision repair to induce programmable, single nucleotide (C→T or A→G) changes in DNA without generating double-strand DNA breaks, without requiring a donor DNA template, and without inducing an excess of stochastic insertions and deletions.

Details of nucleobase editing proteins are described in International PCT Application Nos. PCT/2017/045381 (WO2018/027078) and PCT/US2016/058344 (WO2017/070632), each of which is incorporated herein by reference for its entirety. Also see Komor, A.C., *et al.*, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., *et al.*, "Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); and Komor, A.C., *et al.*, "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017), the entire contents of which are hereby incorporated by reference.

Use of a self-inactivating base editor system provided herein comprises the steps of: (a) contacting a target nucleotide sequence of a polynucleotide (*e.g.*, double- or single stranded DNA or RNA) of a subject with a base editor system comprising a nucleobase editor (*e.g.*, an adenosine base editor or a cytidine base editor) and a guide polynucleic acid (*e.g.*, gRNA), wherein the target nucleotide sequence comprises a targeted nucleobase pair; (b) inducing strand separation of said target region; (c) converting a first nucleobase of said target nucleobase pair in a single strand of the target region to a second nucleobase; (d) cutting no more than one strand of said target region, where a third nucleobase complementary to the first nucleobase base is replaced by a fourth nucleobase complementary to the second nucleobase; (e) contacting a target intron sequence present in an open reading frame encoding a domain of the nucleobase editor with a guide RNA that targets a splice acceptor or splice donor site of the intron and introducing an edit as described in steps b-d, thereby inactivating the base editor. Inactivation can be induced at any time when a desired level of editing has been reached. It should be appreciated that in some embodiments, step (b) or (e) is omitted. In some embodiments, said targeted nucleobase pair is a plurality of nucleobase pairs in one or more genes. In some embodiments, the base editor

system provided herein is capable of multiplex editing of a plurality of nucleobase pairs in one or more genes. In some embodiments, the plurality of nucleobase pairs is located in the same gene. In some embodiments, the plurality of nucleobase pairs is located in one or more genes, wherein at least one gene is located in a different locus.

In some embodiments, the cut single strand (nicked strand) is hybridized to the guide nucleic acid. In some embodiments, the cut single strand is opposite to the strand comprising the first nucleobase. In some embodiments, the base editor comprises a Cas9 domain. In some embodiments, the first base is adenine, and the second base is not a G, C, A, or T. In some embodiments, the second base is inosine.

In some embodiments, a single guide polynucleotide may be utilized to target a deaminase to a target nucleic acid sequence. In some embodiments, a single pair of guide polynucleotides may be utilized to target different deaminases to a target nucleic acid sequence.

The components of a base editor system (e.g., a deaminase domain, a guide RNA, and/or a polynucleotide programmable nucleotide binding domain) may be associated with each other covalently or non-covalently. For example, in some embodiments, the deaminase domain can be targeted to a target nucleotide sequence by a polynucleotide programmable nucleotide binding domain, optionally where the polynucleotide programmable nucleotide binding domain is complexed with a polynucleotide (e.g., a guide RNA). In some embodiments, a polynucleotide programmable nucleotide binding domain can be fused or linked to a deaminase domain. In some embodiments, a polynucleotide programmable nucleotide binding domain can target a deaminase domain to a target nucleotide sequence by non-covalently interacting with or associating with the deaminase domain. For example, in some embodiments, the nucleobase editing component (*e.g.*, the deaminase component) comprises an additional heterologous portion or domain that is capable of interacting with, associating with, or capable of forming a complex with a corresponding heterologous portion, antigen, or domain that is part of a polynucleotide programmable nucleotide binding domain and/or a guide polynucleotide (e.g., a guide RNA) complexed therewith. In some embodiments, the polynucleotide programmable nucleotide binding domain, and/or a guide polynucleotide (e.g., a guide RNA) complexed therewith, comprises an additional heterologous portion or domain that is capable of interacting with, associating with, or capable of forming a complex with a corresponding heterologous portion, antigen, or domain that is part of a nucleobase editing domain (e.g., the deaminase component). In some embodiments, the additional heterologous portion may be capable of binding to, interacting

with, associating with, or forming a complex with a polypeptide. In some embodiments, the additional heterologous portion may be capable of binding to, interacting with, associating with, or forming a complex with a polynucleotide. In some embodiments, the additional heterologous portion may be capable of binding to a guide polynucleotide. In some embodiments, the additional heterologous portion may be capable of binding to a polypeptide linker. In some embodiments, the additional heterologous portion is capable of binding to a polynucleotide linker. An additional heterologous portion may be a protein domain. In some embodiments, an additional heterologous portion comprises a polypeptide, such as a 22 amino acid RNA-binding domain of the lambda bacteriophage antiterminator protein N (N22p), a 2G12 IgG homodimer domain, an ABI, an antibody (*e.g.* an antibody that binds a component of the base editor system or a heterologous portion thereof) or fragment thereof (*e.g.* heavy chain domain 2 (CH2) of IgM (MHD2) or IgE (EHD2), an immunoglobulin Fc region, a heavy chain domain 3 (CH3) of IgG or IgA, a heavy chain domain 4 (CH4) of IgM or IgE, an Fab, an Fab2, miniantibodies, and/or ZIP antibodies), a barnase-barstar dimer domain, a Bcl-xL domain, a Calcineurin A (CAN) domain, a Cardiac phospholamban transmembrane pentamer domain, a collagen domain, a Com RNA binding protein domain (*e.g.* SfMu Com coat protein domain, and SfMu Com binding protein domain), a Cyclophilin-Fas fusion protein (CyP-Fas) domain, a Fab domain, an Fe domain, a fibritin foldon domain, an FK506 binding protein (FKBP) domain, an FKBP binding domain (FRB) domain of mTOR, a foldon domain, a fragment X domain, a GAI domain, a GID1 domain, a Glycophorin A transmembrane domain, a GyrB domain, a Halo tag, an HIV Gp41 trimerisation domain, an HPV45 oncoprotein E7 C-terminal dimer domain, a hydrophobic polypeptide, a K Homology (KH) domain, a Ku protein domain (e.g., a Ku heterodimer), a leucine zipper, a LOV domain, a mitochondrial antiviral-signaling protein CARD filament domain, an MS2 coat protein domain (MCP), a non-natural RNA aptamer ligand that binds a corresponding RNA motif/aptamer, a parathyroid hormone dimerization domain, a PP7 coat protein (PCP) domain, a PSD95-Dlgl-zo-1 (PDZ) domain, a PYL domain, a SNAP tag, a SpyCatcher moiety, a SpyTag moiety, a streptavidin domain, a streptavidin-binding protein domain, a streptavidin binding protein (SBP) domain, a telomerase Sm7 protein domain (*e.g.* Sm7 homoheptamer or a monomeric Sm-like protein), and/or fragments thereof. In embodiments, an additional heterologous portion comprises a polynucleotide (e.g., an RNA motif), such as an MS2 phage operator stem-loop (*e.g.* an MS2, an MS2 C-5 mutant, or an MS2 F-5 mutant), a non-natural RNA motif, a PP7 opterator stem-loop, an SfMu phate Com stem-loop, a steril alpha motif, a telomerase Ku binding motif, a telomerase Sm7 binding

motif, and/or fragments thereof. Non-limiting examples of additional heterologous portions include polypeptides with at least about 85% sequence identity to any one or more of SEQ ID NOs: 492, 494, 496, 498-500, or fragments thereof. Non-limiting examples of additional heterologous portions include polynucleotides with at least about 85% sequence identity to any one or more of SEQ ID NOs: 491, 493, 495, 497 or fragments thereof.

A base editor system may further comprise a guide polynucleotide component. It should be appreciated that components of the base editor system may be associated with each other via covalent bonds, noncovalent interactions, or any combination of associations and interactions thereof. In some embodiments, a deaminase domain can be targeted to a target nucleotide sequence by a guide polynucleotide. For example, in some embodiments, the nucleobase editing component of the base editor system (*e.g.*, the deaminase component) comprises an additional heterologous portion or domain (*e.g.*, polynucleotide binding domain such as an RNA or DNA binding protein) that is capable of interacting with, associating with, or capable of forming a complex with a heterologous portion or segment (*e.g.*, a polynucleotide motif), or antigen of a guide polynucleotide. In some embodiments, the additional heterologous portion or domain (*e.g.*, polynucleotide binding domain such as an RNA or DNA binding protein) can be fused or linked to the deaminase domain. In some embodiments, the additional heterologous portion may be capable of binding to, interacting with, associating with, or forming a complex with a polypeptide. In some embodiments, the additional heterologous portion may be capable of binding to, interacting with, associating with, or forming a complex with a polynucleotide. In some embodiments, the additional heterologous portion may be capable of binding to a guide polynucleotide. In some embodiments, the additional heterologous portion may be capable of binding to a polypeptide linker. In some embodiments, the additional heterologous portion may be capable of binding to a polynucleotide linker. An additional heterologous portion may be a protein domain. In some embodiments, an additional heterologous portion comprises a polypeptide, such as a 22 amino acid RNA-binding domain of the lambda bacteriophage antiterminator protein N (N22p), a 2G12 IgG homodimer domain, an ABI, an antibody (*e.g.* an antibody that binds a component of the base editor system or a heterologous portion thereof) or fragment thereof (*e.g.* heavy chain domain 2 (CH2) of IgM (MHD2) or IgE (EHD2), an immunoglobulin Fc region, a heavy chain domain 3 (CH3) of IgG or IgA, a heavy chain domain 4 (CH4) of IgM or IgE, an Fab, an Fab2, miniantibodies, and/or ZIP antibodies), a barnase-barstar dimer domain, a Bcl-xL domain, a Calcineurin A (CAN) domain, a Cardiac phospholamban transmembrane pentamer domain, a collagen domain, a Com RNA binding protein domain

(*e.g.* SfMu Com coat protein domain, and SfMu Com binding protein domain), a Cyclophilin-Fas fusion protein (CyP-Fas) domain, a Fab domain, an Fe domain, a fibritin foldon domain, an FK506 binding protein (FKBP) domain, an FKBP binding domain (FRB) domain of mTOR, a foldon domain, a fragment X domain, a GAI domain, a GID1 domain, a Glycophorin A transmembrane domain, a GyrB domain, a Halo tag, an HIV Gp41 trimerisation domain, an HPV45 oncoprotein E7 C-terminal dimer domain, a hydrophobic polypeptide, a K Homology (KH) domain, a Ku protein domain (e.g., a Ku heterodimer), a leucine zipper, a LOV domain, a mitochondrial antiviral-signaling protein CARD filament domain, an MS2 coat protein domain (MCP), a non-natural RNA aptamer ligand that binds a corresponding RNA motif/aptamer, a parathyroid hormone dimerization domain, a PP7 coat protein (PCP) domain, a PSD95-Dlgl-zo-1 (PDZ) domain, a PYL domain, a SNAP tag, a SpyCatcher moiety, a SpyTag moiety, a streptavidin domain, a streptavidin-binding protein domain, a streptavidin binding protein (SBP) domain, a telomerase Sm7 protein domain (*e.g.* Sm7 homoheptamer or a monomeric Sm-like protein), and/or fragments thereof. In embodiments, an additional heterologous portion comprises a polynucleotide (e.g., an RNA motif), such as an MS2 phage operator stem-loop (*e.g.* an MS2, an MS2 C-5 mutant, or an MS2 F-5 mutant), a non-natural RNA motif, a PP7 opterator stem-loop, an SfMu phate Com stem-loop, a steril alpha motif, a telomerase Ku binding motif, a telomerase Sm7 binding motif, and/or fragments thereof. Non-limiting examples of additional heterologous portions include polypeptides with at least about 85% sequence identity to any one or more of SEQ ID NOs: 492, 494, 496, 498-500, or fragments thereof. Non-limiting examples of additional heterologous portions include polynucleotides with at least about 85% sequence identity to any one or more of SEQ ID NOs: 491, 493, 495, 497 or fragments thereof..

In some embodiments, a base editor system can further comprise an inhibitor of base excision repair (BER) component. It should be appreciated that components of the base editor system may be associated with each other via covalent bonds, noncovalent interactions, or any combination of associations and interactions thereof. The inhibitor of BER component may comprise a base excision repair inhibitor. In some embodiments, the inhibitor of base excision repair can be a uracil DNA glycosylase inhibitor (UGI). In some embodiments, the inhibitor of base excision repair can be an inosine base excision repair inhibitor. In some embodiments, the inhibitor of base excision repair can be targeted to the target nucleotide sequence by the polynucleotide programmable nucleotide binding domain, optionally where the polynucleotide programmable nucleotide binding domain is complexed with a polynucleotide (e.g., a guide RNA). In some embodiments, a polynucleotide programmable

nucleotide binding domain can be fused or linked to an inhibitor of base excision repair. In some embodiments, a polynucleotide programmable nucleotide binding domain can be fused or linked to a deaminase domain and an inhibitor of base excision repair. In some embodiments, a polynucleotide programmable nucleotide binding domain can target an inhibitor of base excision repair to a target nucleotide sequence by non-covalently interacting with or associating with the inhibitor of base excision repair. For example, in some embodiments, the inhibitor of base excision repair component comprises an additional heterologous portion or domain that is capable of interacting with, associating with, or capable of forming a complex with a corresponding additional heterologous portion, antigen, or domain that is part of a polynucleotide programmable nucleotide binding domain. In some embodiments, the polynucleotide programming nucleotide binding domain component, and/or a guide polynucleotide (e.g., a guide RNA) complexed therewith, comprises an additional heterologous portion or domain that is capable of interacting with, associating with, or capable of forming a corresponding heterologous portion, antigen, or domain that is part of an inhibitor of base excision repair component. In some embodiments, the inhibitor of base excision repair can be targeted to the target nucleotide sequence by the guide polynucleotide. For example, in some embodiments, the inhibitor of base excision repair comprises an additional heterologous portion or domain (*e.g.*, polynucleotide binding domain such as an RNA or DNA binding protein) that is capable of interacting with, associating with, or capable of forming a complex with a portion or segment (*e.g.*, a polynucleotide motif) of a guide polynucleotide. In some embodiments, the additional heterologous portion or domain of the guide polynucleotide (*e.g.*, polynucleotide binding domain such as an RNA or DNA binding protein) can be fused or linked to the inhibitor of base excision repair. In some embodiments, the additional heterologous portion may be capable of binding to, interacting with, associating with, or forming a complex with a polynucleotide. In some embodiments, the additional heterologous portion may be capable of binding to a guide polynucleotide. In some embodiments, the additional heterologous portion may be capable of binding to a polypeptide linker. In some embodiments, the additional heterologous portion may be capable of binding to a polynucleotide linker. An additional heterologous portion may be a protein domain. In some embodiments, an additional heterologous portion comprises a polypeptide, such as a 22 amino acid RNA-binding domain of the lambda bacteriophage antiterminator protein N (N22p), a 2G12 IgG homodimer domain, an ABI, an antibody (*e.g.* an antibody that binds a component of the base editor system or a heterologous portion thereof) or fragment thereof (*e.g.* heavy chain domain 2 (CH2) of IgM (MHD2) or IgE

(EHD2), an immunoglobulin Fc region, a heavy chain domain 3 (CH3) of IgG or IgA, a heavy chain domain 4 (CH4) of IgM or IgE, an Fab, an Fab2, miniantibodies, and/or ZIP antibodies), a barnase-barstar dimer domain, a Bcl-xL domain, a Calcineurin A (CAN) domain, a Cardiac phospholamban transmembrane pentamer domain, a collagen domain, a Com RNA binding protein domain (*e.g.* SfMu Com coat protein domain, and SfMu Com binding protein domain), a Cyclophilin-Fas fusion protein (CyP-Fas) domain, a Fab domain, an Fe domain, a fibritin foldon domain, an FK506 binding protein (FKBP) domain, an FKBP binding domain (FRB) domain of mTOR, a foldon domain, a fragment X domain, a GAI domain, a GID1 domain, a Glycophorin A transmembrane domain, a GyrB domain, a Halo tag, an HIV Gp41 trimerisation domain, an HPV45 oncoprotein E7 C-terminal dimer domain, a hydrophobic polypeptide, a K Homology (KH) domain, a Ku protein domain (e.g., a Ku heterodimer), a leucine zipper, a LOV domain, a mitochondrial antiviral-signaling protein CARD filament domain, an MS2 coat protein domain (MCP), a non-natural RNA aptamer ligand that binds a corresponding RNA motif/aptamer, a parathyroid hormone dimerization domain, a PP7 coat protein (PCP) domain, a PSD95-Dlgl-zo-1 (PDZ) domain, a PYL domain, a SNAP tag, a SpyCatcher moiety, a SpyTag moiety, a streptavidin domain, a streptavidin-binding protein domain, a streptavidin binding protein (SBP) domain, a telomerase Sm7 protein domain (*e.g.* Sm7 homoheptamer or a monomeric Sm-like protein), and/or fragments thereof. In embodiments, an additional heterologous portion comprises a polynucleotide (e.g., an RNA motif), such as an MS2 phage operator stem-loop (*e.g.* an MS2, an MS2 C-5 mutant, or an MS2 F-5 mutant), a non-natural RNA motif, a PP7 opterator stem-loop, an SfMu phate Com stem-loop, a steril alpha motif, a telomerase Ku binding motif, a telomerase Sm7 binding motif, and/or fragments thereof. Non-limiting examples of additional heterologous portions include polypeptides with at least about 85% sequence identity to any one or more of SEQ ID NOs: 492, 494, 496, 498-500, or fragments thereof. Non-limiting examples of additional heterologous portions include polynucleotides with at least about 85% sequence identity to any one or more of SEQ ID NOs: 491, 493, 495, 497, or fragments thereof.

In some instances, components of the base editing system are associated with one another through the interaction of leucine zipper domains (e.g., SEQ ID NOs: 499 and 500). In some cases, components of the base editing system are associated with one another through polypeptide domains (e.g., FokI domains) that associate to form protein complexes containing about, at least about, or no more than about 1, 2 (i.e., dimerize), 3, 4, 5, 6, 7, 8, 9,

10 polypeptide domain units, optionally the polypeptide domains may include alterations that reduce or eliminate an activity thereof.

In some instances, components of the base editing system are associated with one another through the interaction of multimeric antibodies or fragments thereof (e.g., IgG, IgD, IgA, IgM, IgE, a heavy chain domain 2 (CH2) of IgM (MHD2) or IgE (EHD2), an immunoglobulin Fc region, a heavy chain domain 3 (CH3) of IgG or IgA, a heavy chain domain 4 (CH4) of IgM or IgE, an Fab, and an Fab2). In some instances, the antibodies are dimeric, trimeric, or tetrameric. In embodiments, the dimeric antibodies bind a polypeptide or polynucleotide component of the base editing system.

In some cases, components of the base editing system are associated with one another through the interaction of a polynucleotide-binding protein domain(s) with a polynucleotide(s). In some instances, components of the base editing system are associated with one another through the interaction of one or more polynucleotide-binding protein domains with polynucleotides that are self complementary and/or complementary to one another so that complementary binding of the polynucleotides to one another brings into association their respective bound polynucleotide-binding protein domain(s).

In some instances, components of the base editing system are associated with one another through the interaction of a polypeptide domain(s) with a small molecule(s) (e.g., chemical inducers of dimerization (CIDs), also known as "dimerizers"). Non-limiting examples of CIDs include those disclosed in Amara, *et al.*, "A versatile synthetic dimerizer for the regulation of protein-protein interactions," *PNAS*, 94:10618-10623 (1997); and Voß, *et al.* "Chemically induced dimerization: reversible and spatiotemporal control of protein function in cells," *Current Opinion in Chemical Biology*, 28:194-201 (2015), the disclosures of each of which are incorporated herein by reference in their entireties for all purposes. Non-limiting examples of polypeptides that can dimerize and their corresponding dimerizing agents are provided in **Table 10.1** below.

Table 10.1. Chemically induced dimerization systems.

Dimerizing Polypeptides		**Dimerizing agent**
FKBP	FKBP	FK1012
FKBP	Calcineurin A (CNA)	FK506

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	PYRMandi	Mandipropamid
SNAP-tag	HaloTag	HaXS
eDHFR	HaloTag	TMP-HTag
Bcl-xL	Fab (AZ1)	ABT-737

In embodiments, the additional heterologous portion is part of a guide RNA molecule. In some instances, the additional heterologous portion contains or is an RNA motif. The RNA motif may be positioned at the 5′ or 3′ end of the guide RNA molecule or various positions of a guide RNA molecule. In embodiments, the RNA motif is positioned within the guide RNA to reduce steric hindrance, optionally where such hindrance is associated with other bulky loops of an RNA scaffold. In some instances, it is advantageous to link the RNA motif is linked to other portions of the guide RNA by way of a linker, where the linker can be about, at least about, or no more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more nucleotides in length. Optionally, the linker contains a GC-rich nucleotide sequence. The guide RNA can contain 1, 2, 3, 4, 5, or more copies of the RNA motif, optionally where they are positioned consecutively, and/or optionally where they are each separated from one another by a linker(s). The RNA motif may include any one or more of the polynucleotide modifications described herein. Non-limiting examples of suitable modifications to the RNA motif include 2' deoxy-2-aminopurine, 2'ribose-2-aminopurine, phosphorothioate mods, 2'-Omethyl mods, 2'-Fluro mods and LNA mods. Advantageously, the modifications help to increase stability and promote stronger bonds/folding structure of a hairpin(s) formed by the RNA motif.

In some embodiments, the RNA motif is modified to include an extension. In embodiments, the extension contains about, at least about, or no more than about 2, 3, 4, 5, 10, 15, 20, or 25 nucleotides. In some instances, the extension results in an alteration in the

length of a stem formed by the RNA motif (e.g., a lengthening or a shortening). It can be advantageous for a stem formed by the RNA motif to be about, at least about, or no more than about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 nucleotides in length. In various embodiments, the extension increases flexibility of the RNA motif and/or increases binding with a corresponding RNA motif.

In some embodiments, the base editor inhibits base excision repair (BER) of the edited strand. In some embodiments, the base editor protects or binds the non-edited strand. In some embodiments, the base editor comprises UGI activity. In some embodiments, the base editor comprises a catalytically inactive inosine-specific nuclease. In some embodiments, the base editor comprises nickase activity. In some embodiments, the intended edit of base pair is upstream of a PAM site. In some embodiments, the intended edit of base pair is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides upstream of the PAM site. In some embodiments, the intended edit of base-pair is downstream of a PAM site. In some embodiments, the intended edited base pair is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides downstream stream of the PAM site.

In some embodiments, the method does not require a canonical (*e.g.*, NGG) PAM site. In some embodiments, the nucleobase editor comprises a linker or a spacer. In some embodiments, the linker or spacer is 1-25 amino acids in length. In some embodiments, the linker or spacer is 5-20 amino acids in length. In some embodiments, the linker or spacer is 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 amino acids in length.

In some embodiments, the base editing fusion proteins provided herein need to be positioned at a precise location, for example, where a target base is placed within a defined region (*e.g.*, a "deamination window"). In some embodiments, a target can be within a 4 base region. In some embodiments, such a defined target region can be approximately 15 bases upstream of the PAM. See Komor, A.C., *et al.*, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., *et al.*, "Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); and Komor, A.C., *et al.*, "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017), the entire contents of which are hereby incorporated by reference.

In some embodiments, the target region comprises a target window, wherein the target window comprises the target nucleobase pair. In some embodiments, the target window comprises 1- 10 nucleotides. In some embodiments, the target window is 1, 2, 3, 4, 5, 6, 7, 8,

9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides in length. In some embodiments, the intended edit of base pair is within the target window. In some embodiments, the target window comprises the intended edit of base pair. In some embodiments, the method is performed using any of the base editors provided herein. In some embodiments, a target window is a deamination window. A deamination window can be the defined region in which a base editor acts upon and deaminates a target nucleotide. In some embodiments, the deamination window is within a 2, 3, 4, 5, 6, 7, 8, 9, or 10 base regions. In some embodiments, the deamination window is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 bases upstream of the PAM.

The base editors of the present disclosure can comprise any domain, feature or amino acid sequence which facilitates the editing of a target polynucleotide sequence. For example, in some embodiments, the base editor comprises a nuclear localization sequence (NLS). In some embodiments, an NLS of the base editor is localized between a deaminase domain and a polynucleotide programmable nucleotide binding domain. In some embodiments, an NLS of the base editor is localized C-terminal to a polynucleotide programmable nucleotide binding domain.

Other exemplary features that can be present in a base editor as disclosed herein are localization sequences, such as cytoplasmic localization sequences, export sequences, such as nuclear export sequences, or other localization sequences, as well as sequence tags that are useful for solubilization, purification, or detection of the fusion proteins. Suitable protein tags provided herein include, but are not limited to, biotin carboxylase carrier protein (BCCP) tags, myc-tags, calmodulin-tags, FLAG-tags, hemagglutinin (HA)-tags, polyhistidine tags, also referred to as histidine tags or His-tags, maltose binding protein (MBP)-tags, nus-tags, glutathione-S-transferase (GST)-tags, green fluorescent protein (GFP)-tags, thioredoxin-tags, S-tags, Softags (*e.g.*, Softag 1, Softag 3), strep-tags, biotin ligase tags, FlAsH tags, V5 tags, and SBP-tags. Additional suitable sequences will be apparent to those of skill in the art. In some embodiments, the fusion protein comprises one or more His tags.

In some embodiments, non-limiting exemplary cytidine base editors (CBE) include BE1 (APOBEC1-XTEN-dCas9), BE2 (APOBEC1-XTEN-dCas9-UGI), BE3 (APOBEC1-XTEN-dCas9(A840H)-UGI), BE3-Gam, saBE3, saBE4-Gam, BE4, BE4-Gam, saBE4, or saB4E-Gam. BE4 extends the APOBEC1-Cas9n(D10A) linker to 32 amino acids and the Cas9n-UGI linker to 9 amino acids, and appends a second copy of UGI to the C-terminus of the construct with another 9-amino acid linker into a single base editor construct. The base editors saBE3 and saBE4 have the *S. pyogenes* Cas9n(D10A) replaced with the smaller *S.*

aureus Cas9n(D10A). BE3-Gam, saBE3-Gam, BE4-Gam, and saBE4-Gam have 174 residues of Gam protein fused to the N-terminus of BE3, saBE3, BE4, and saBE4 via the 16 amino acid XTEN linker.

In some embodiments, the adenosine base editor (ABE) can deaminate adenine in DNA. In some embodiments, ABE is generated by replacing APOBEC1 component of BE3 with natural or engineered *E. coli* TadA, human ADAR2, mouse ADA, or human ADAT2. In some embodiments, ABE comprises evolved TadA variant. In some embodiments, the ABE is ABE 1.2 (TadA*-XTEN-nCas9-NLS). In some embodiments, TadA* comprises A106V and D108N mutations.

In some embodiments, the ABE is a second-generation ABE. In some embodiments, the ABE is ABE2.1, which comprises additional mutations D147Y and E155V in TadA* (TadA*2.1). In some embodiments, the ABE is ABE2.2, ABE2.1 fused to catalytically inactivated version of human alkyl adenine DNA glycosylase (AAG with E125Q mutation). In some embodiments, the ABE is ABE2.3, ABE2.1 fused to catalytically inactivated version of *E. coli* Endo V (inactivated with D35A mutation). In some embodiments, the ABE is ABE2.6 which has a linker twice as long (32 amino acids, $(SGGS)_2$ (SEQ ID NO: 418)-XTEN-$(SGGS)_2$ (SEQ ID NO: 418)) as the linker in ABE2.1. In some embodiments, the ABE is ABE2.7, which is ABE2.1 tethered with an additional wild-type TadA monomer. In some embodiments, the ABE is ABE2.8, which is ABE2.1 tethered with an additional TadA*2.1 monomer. In some embodiments, the ABE is ABE2.9, which is a direct fusion of evolved TadA (TadA*2.1) to the N-terminus of ABE2.1. In some embodiments, the ABE is ABE2.10, which is a direct fusion of wild-type TadA to the N-terminus of ABE2.1. In some embodiments, the ABE is ABE2.11, which is ABE2.9 with an inactivating E59A mutation at the N-terminus of TadA* monomer. In some embodiments, the ABE is ABE2.12, which is ABE2.9 with an inactivating E59A mutation in the internal TadA* monomer.

In some embodiments, the ABE is a third generation ABE. In some embodiments, the ABE is ABE3.1, which is ABE2.3 with three additional TadA mutations (L84F, H123Y, and I156F).

In some embodiments, the ABE is a fourth generation ABE. In some embodiments, the ABE is ABE4.3, which is ABE3.1 with an additional TadA mutation A142N (TadA*4.3).

In some embodiments, the ABE is a fifth generation ABE. In some embodiments, the ABE is ABE5.1, which is generated by importing a consensus set of mutations from surviving clones (H36L, R51L, S146C, and K157N) into ABE3.1. In some embodiments, the ABE is ABE5.3, which has a heterodimeric construct containing wild-type *E. coli* TadA

fused to an internal evolved TadA*. In some embodiments, the ABE is ABE5.2, ABE5.4, ABE5.5, ABE5.6, ABE5.7, ABE5.8, ABE5.9, ABE5.10, ABE5.11, ABE5.12, ABE5.13, or ABE5.14, as shown in Table 10 below. In some embodiments, the ABE is a sixth generation ABE. In some embodiments, the ABE is ABE6.1, ABE6.2, ABE6.3, ABE6.4, ABE6.5, or ABE6.6, as shown in Table 10 below. In some embodiments, the ABE is a seventh generation ABE. In some embodiments, the ABE is ABE7.1, ABE7.2, ABE7.3, ABE7.4, ABE7.5, ABE7.6, ABE7.7, ABE7.8, ABE 7.9, or ABE7.10, as shown in Table 10 below.

Table 10. Genotypes of ABEs

	23	26	36	37	48	49	51	72	84	87	106	108	123	125	142	146	147	152	155	156	157	161
ABE0.1	W	R	H	N	P		R	N	L	S	A	D	H	G	A	S	D	R	E	I	K	K
ABE0.2	W	R	H	N	P		R	N	L	S	A	D	H	G	A	S	D	R	E	I	K	K
ABE1.1	W	R	H	N	P		R	N	L	S	A	N	H	G	A	S	D	R	E	I	K	K
ABE1.2	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	D	R	E	I	K	K
ABE2.1	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.2	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.3	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.4	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.5	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.6	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.7	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.8	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.9	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.10	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.11	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.12	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE3.1	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.2	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.3	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.4	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.5	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.6	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.7	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.8	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE4.1	W	R	H	N	P		R	N	L	S	V	N	H	G	N	S	Y	R	V	I	K	K
ABE4.2	W	G	H	N	P		R	N	L	S	V	N	H	G	N	S	Y	R	V	I	K	K
ABE4.3	W	R	H	N	P		R	N	F	S	V	N	Y	G	N	S	Y	R	V	F	K	K
ABE5.1	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.2	W	R	H	S	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	T
ABE5.3	W	R	L	N	P		L	N	I	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.4	W	R	H	S	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	T
ABE5.5	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.6	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.7	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.8	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.9	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K

	23	26	36	37	48	49	51	72	84	87	106	108	123	125	142	146	147	152	155	156	157	161
ABE5.10	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.11	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.12	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.13	W	R	H	N	P		L	D	F	S	V	N	Y	A	A	S	Y	R	V	F	K	K
ABE5.14	W	R	H	N	S		L	N	F	C	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE6.1	W	R	H	N	S		L	N	F	S	V	N	Y	G	N	S	Y	R	V	F	K	K
ABE6.2	W	R	H	N	T	V	L	N	F	S	V	N	Y	G	N	S	Y	R	V	F	N	K
ABE6.3	W	R	L	N	S		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE6.4	W	R	L	N	S		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE6.5	W	R	L	N	T	V	L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE6.6	W	R	L	N	T	V	L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE7.1	W	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE7.2	W	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE7.3	L	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE7.4	R	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE7.5	W	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	H	V	F	N	K
ABE7.6	W	R	L	N	A		L	N	I	S	V	N	Y	G	A	C	Y	P	V	F	N	K
ABE7.7	L	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K
ABE7.8	L	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE7.9	L	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	P	V	F	N	K
ABE7.10	R	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K

In some embodiments, the base editor is an eighth generation ABE (ABE8). In some embodiments, the ABE8 contains a TadA*8 variant. In some embodiments, the ABE8 has a monomeric construct containing a TadA*8 variant ("ABE8.x-m"). In some embodiments, the ABE8 is ABE8.1-m, which has a monomeric construct containing TadA*7.10 with a Y147T mutation (TadA*8.1). In some embodiments, the ABE8 is ABE8.2-m, which has a monomeric construct containing TadA*7.10 with a Y147R mutation (TadA*8.2). In some embodiments, the ABE8 is ABE8.3-m, which has a monomeric construct containing TadA*7.10 with a Q154S mutation (TadA*8.3). In some embodiments, the ABE8 is ABE8.4-m, which has a monomeric construct containing TadA*7.10 with a Y123H mutation (TadA*8.4). In some embodiments, the ABE8 is ABE8.5-m, which has a monomeric construct containing TadA*7.10 with a V82S mutation (TadA*8.5). In some embodiments, the ABE8 is ABE8.6-m, which has a monomeric construct containing TadA*7.10 with a T166R mutation (TadA*8.6). In some embodiments, the ABE8 is ABE8.7-m, which has a monomeric construct containing TadA*7.10 with a Q154R mutation (TadA*8.7). In some embodiments, the ABE8 is ABE8.8-m, which has a monomeric construct containing TadA*7.10 with Y147R, Q154R, and Y123H mutations (TadA*8.8). In some embodiments, the ABE8 is ABE8.9-m, which has a monomeric construct containing TadA*7.10 with

Y147R, Q154R and I76Y mutations (TadA*8.9). In some embodiments, the ABE8 is ABE8.10-m, which has a monomeric construct containing TadA*7.10 with Y147R, Q154R, and T166R mutations (TadA*8.10). In some embodiments, the ABE8 is ABE8.11-m, which has a monomeric construct containing TadA*7.10 with Y147T and Q154R mutations (TadA*8.11). In some embodiments, the ABE8 is ABE8.12-m, which has a monomeric construct containing TadA*7.10 with Y147T and Q154S mutations (TadA*8.12).

In some embodiments, the ABE8 is ABE8.13-m, which has a monomeric construct containing TadA*7.10 with Y123H (Y123H reverted from H123Y), Y147R, Q154R and I76Y mutations (TadA*8.13). In some embodiments, the ABE8 is ABE8.14-m, which has a monomeric construct containing TadA*7.10 with I76Y and V82S mutations (TadA*8.14). In some embodiments, the ABE8 is ABE8.15-m, which has a monomeric construct containing TadA*7.10 with V82S and Y147R mutations (TadA*8.15). In some embodiments, the ABE8 is ABE8.16-m, which has a monomeric construct containing TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y) and Y147R mutations (TadA*8.16). In some embodiments, the ABE8 is ABE8.17-m, which has a monomeric construct containing TadA*7.10 with V82S and Q154R mutations (TadA*8.17). In some embodiments, the ABE8 is ABE8.18-m, which has a monomeric construct containing TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y) and Q154R mutations (TadA*8.18). In some embodiments, the ABE8 is ABE8.19-m, which has a monomeric construct containing TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y), Y147R and Q154R mutations (TadA*8.19). In some embodiments, the ABE8 is ABE8.20-m, which has a monomeric construct containing TadA*7.10 with I76Y, V82S, Y123H (Y123H reverted from H123Y), Y147R and Q154R mutations (TadA*8.20). In some embodiments, the ABE8 is ABE8.21-m, which has a monomeric construct containing TadA*7.10 with Y147R and Q154S mutations (TadA*8.21). In some embodiments, the ABE8 is ABE8.22-m, which has a monomeric construct containing TadA*7.10 with V82S and Q154S mutations (TadA*8.22). In some embodiments, the ABE8 is ABE8.23-m, which has a monomeric construct containing TadA*7.10 with V82S and Y123H (Y123H reverted from H123Y) mutations (TadA*8.23). In some embodiments, the ABE8 is ABE8.24-m, which has a monomeric construct containing TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y), and Y147T mutations (TadA*8.24).

In some embodiments, the ABE8 has a heterodimeric construct containing wild-type *E. coli* TadA fused to a TadA*8 variant ("ABE8.x-d"). In some embodiments, the ABE8 is ABE8.1-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to

TadA*7.10 with a Y147T mutation (TadA*8.1). In some embodiments, the ABE8 is ABE8.2-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with a Y147R mutation (TadA*8.2). In some embodiments, the ABE8 is ABE8.3-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with a Q154S mutation (TadA*8.3). In some embodiments, the ABE8 is ABE8.4-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with a Y123H mutation (TadA*8.4). In some embodiments, the ABE8 is ABE8.5-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with a V82S mutation (TadA*8.5). In some embodiments, the ABE8 is ABE8.6-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with a T166R mutation (TadA*8.6). In some embodiments, the ABE8 is ABE8.7-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with a Q154R mutation (TadA*8.7). In some embodiments, the ABE8 is ABE8.8-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with Y147R, Q154R, and Y123H mutations (TadA*8.8). In some embodiments, the ABE8 is ABE8.9-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with Y147R, Q154R and I76Y mutations (TadA*8.9). In some embodiments, the ABE8 is ABE8.10-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with Y147R, Q154R, and T166R mutations (TadA*8.10). In some embodiments, the ABE8 is ABE8.11-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with Y147T and Q154R mutations (TadA*8.11). In some embodiments, the ABE8 is ABE8.12-d, which has heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with Y147T and Q154S mutations (TadA*8.12). In some embodiments, the ABE8 is ABE8.13-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with Y123H (Y123H reverted from H123Y), Y147R, Q154R and I76Y mutations (TadA*8.13). In some embodiments, the ABE8 is ABE8.14-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with I76Y and V82S mutations (TadA*8.14). In some embodiments, the ABE8 is ABE8.15-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V82S and Y147R mutations (TadA*8.15). In some embodiments, the ABE8 is ABE8.16-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y) and Y147R mutations (TadA*8.16). In some embodiments, the ABE8 is ABE8.17-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V82S

and Q154R mutations (TadA*8.17). In some embodiments, the ABE8 is ABE8.18-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y) and Q154R mutations (TadA*8.18). In some embodiments, the ABE8 is ABE8.19-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y), Y147R and Q154R mutations (TadA*8.19). In some embodiments, the ABE8 is ABE8.20-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with I76Y, V82S, Y123H (Y123H reverted from H123Y), Y147R and Q154R mutations (TadA*8.20). In some embodiments, the ABE8 is ABE8.21-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with Y147R and Q154S mutations (TadA*8.21). In some embodiments, the ABE8 is ABE8.22-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V82S and Q154S mutations (TadA*8.22). In some embodiments, the ABE8 is ABE8.23-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V82S and Y123H (Y123H reverted from H123Y) mutations (TadA*8.23). In some embodiments, the ABE8 is ABE8.24-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y), and Y147T mutations (TadA*8.24).

In some embodiments, the ABE8 has a heterodimeric construct containing TadA*7.10 fused to a TadA*8 variant ("ABE8.x-7"). In some embodiments, the ABE8 is ABE8.1-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with a Y147T mutation (TadA*8.1). In some embodiments, the ABE8 is ABE8.2-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with a Y147R mutation (TadA*8.2). In some embodiments, the ABE8 is ABE8.3-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with a Q154S mutation (TadA*8.3). In some embodiments, the ABE8 is ABE8.4-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with a Y123H mutation (TadA*8.4). In some embodiments, the ABE8 is ABE8.5-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with a V82S mutation (TadA*8.5). In some embodiments, the ABE8 is ABE8.6-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with a T166R mutation (TadA*8.6). In some embodiments, the ABE8 is ABE8.7-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with a Q154R mutation (TadA*8.7). In some embodiments, the ABE8 is ABE8.8-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y147R, Q154R, and Y123H

mutations (TadA*8.8). In some embodiments, the ABE8 is ABE8.9-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y147R, Q154R and I76Y mutations (TadA*8.9). In some embodiments, the ABE8 is ABE8.10-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y147R, Q154R, and T166R mutations (TadA*8.10). In some embodiments, the ABE8 is ABE8.11-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y147T and Q154R mutations (TadA*8.11). In some embodiments, the ABE8 is ABE8.12-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y147T and Q154S mutations (TadA*8.12). In some embodiments, the ABE8 is ABE8.13-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y123H (Y123H reverted from H123Y), Y147R, Q154R and I76Y mutations (TadA*8.13). In some embodiments, the ABE8 is ABE8.14-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with I76Y and V82S mutations (TadA*8.14). In some embodiments, the ABE8 is ABE8.15-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S and Y147R mutations (TadA*8.15). In some embodiments, the ABE8 is ABE8.16-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y) and Y147R mutations (TadA*8.16). In some embodiments, the ABE8 is ABE8.17-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S and Q154R mutations (TadA*8.17). In some embodiments, the ABE8 is ABE8.18-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y) and Q154R mutations (TadA*8.18). In some embodiments, the ABE8 is ABE8.19-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y), Y147R and Q154R mutations (TadA*8.19). In some embodiments, the ABE8 is ABE8.20-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with I76Y, V82S, Y123H (Y123H reverted from H123Y), Y147R and Q154R mutations (TadA*8.20). In some embodiments, the ABE8 is ABE8.21-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y147R and Q154S mutations (TadA*8.21). In some embodiments, the ABE8 is ABE8.22-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S and Q154S mutations (TadA*8.22). In some embodiments, the ABE8 is ABE8.23-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S and Y123H (Y123H reverted from H123Y) mutations (TadA*8.23). In some embodiments, the ABE8 is ABE8.24-7, which has a

heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y), and Y147T mutations (TadA*8.24

In some embodiments, the ABE is ABE8.1-m, ABE8.2-m, ABE8.3-m, ABE8.4-m, ABE8.5-m, ABE8.6-m, ABE8.7-m, ABE8.8-m, ABE8.9-m, ABE8.10-m, ABE8.11-m, ABE8.12-m, ABE8.13-m, ABE8.14-m, ABE8.15-m, ABE8.16-m, ABE8.17-m, ABE8.18-m, ABE8.19-m, ABE8.20-m, ABE8.21-m, ABE8.22-m, ABE8.23-m, ABE8.24-m, ABE8.1-d, ABE8.2-d, ABE8.3-d, ABE8.4-d, ABE8.5-d, ABE8.6-d, ABE8.7-d, ABE8.8-d, ABE8.9-d, ABE8.10-d, ABE8.11-d, ABE8.12-d, ABE8.13-d, ABE8.14-d, ABE8.15-d, ABE8.16-d, ABE8.17-d, ABE8.18-d, ABE8.19-d, ABE8.20-d, ABE8.21-d, ABE8.22-d, ABE8.23-d, or ABE8.24-d as shown in Table 11 below.

Table 11: Adenosine Base Editor 8 (ABE8) Variants

ABE8	Adenosine Deaminase	Adenosine Deaminase Description
ABE8.1-m	TadA*8.1	Monomer_TadA*7.10 + Y147T
ABE8.2-m	TadA*8.2	Monomer_TadA*7.10 + Y147R
ABE8.3-m	TadA*8.3	Monomer_TadA*7.10 + Q154S
ABE8.4-m	TadA*8.4	Monomer_TadA*7.10 + Y123H
ABE8.5-m	TadA*8.5	Monomer_TadA*7.10 + V82S
ABE8.6-m	TadA*8.6	Monomer_TadA*7.10 + T166R
ABE8.7-m	TadA*8.7	Monomer_TadA*7.10 + Q154R
ABE8.8-m	TadA*8.8	Monomer_TadA*7.10 + Y147R_Q154R_Y123H
ABE8.9-m	TadA*8.9	Monomer_TadA*7.10 + Y147R_Q154R_I76Y
ABE8.10-m	TadA*8.10	Monomer_TadA*7.10 + Y147R_Q154R_T166R
ABE8.11-m	TadA*8.11	Monomer_TadA*7.10 + Y147T_Q154R
ABE8.12-m	TadA*8.12	Monomer_TadA*7.10 + Y147T_Q154S
ABE8.13-m	TadA*8.13	Monomer_TadA*7.10 + Y123H_Y147R_Q154R_I76Y
ABE8.14-m	TadA*8.14	Monomer_TadA*7.10 + I76Y_V82S
ABE8.15-m	TadA*8.15	Monomer_TadA*7.10 + V82S_Y147R
ABE8.16-m	TadA*8.16	Monomer_TadA*7.10 + V82S_Y123H_Y147R
ABE8.17-m	TadA*8.17	Monomer_TadA*7.10 + V82S_Q154R
ABE8.18-m	TadA*8.18	Monomer_TadA*7.10 + V82S_Y123H_Q154R
ABE8.19-m	TadA*8.19	Monomer_TadA*7.10 + V82S_Y123H_Y147R_Q154R
ABE8.20-m	TadA*8.20	Monomer_TadA*7.10 + I76Y_V82S_Y123H_Y147R_Q154R
ABE8.21-m	TadA*8.21	Monomer_TadA*7.10 + Y147R_Q154S
ABE8.22-m	TadA*8.22	Monomer_TadA*7.10 + V82S_Q154S
ABE8.23-m	TadA*8.23	Monomer_TadA*7.10 + V82S_Y123H
ABE8.24-m	TadA*8.24	Monomer_TadA*7.10 + V82S_Y123H_Y147T
ABE8.1-d	TadA*8.1	Heterodimer_(WT) + (TadA*7.10 + Y147T)

ABE8.2-d	TadA*8.2	Heterodimer_(WT) + (TadA*7.10 + Y147R)
ABE8.3-d	TadA*8.3	Heterodimer_(WT) + (TadA*7.10 + Q154S)
ABE8.4-d	TadA*8.4	Heterodimer_(WT) + (TadA*7.10 + Y123H)
ABE8.5-d	TadA*8.5	Heterodimer_(WT) + (TadA*7.10 + V82S)
ABE8.6-d	TadA*8.6	Heterodimer_(WT) + (TadA*7.10 + T166R)
ABE8.7-d	TadA*8.7	Heterodimer_(WT) + (TadA*7.10 + Q154R)
ABE8.8-d	TadA*8.8	Heterodimer_(WT) + (TadA*7.10 + Y147R_Q154R_Y123H)
ABE8.9-d	TadA*8.9	Heterodimer_(WT) + (TadA*7.10 + Y147R_Q154R_I76Y)
ABE8.10-d	TadA*8.10	Heterodimer_(WT) + (TadA*7.10 + Y147R_Q154R_T166R)
ABE8.11-d	TadA*8.11	Heterodimer_(WT) + (TadA*7.10 + Y147T_Q154R)
ABE8.12-d	TadA*8.12	Heterodimer_(WT) + (TadA*7.10 + Y147T_Q154S)
ABE8.13-d	TadA*8.13	Heterodimer_(WT) + (TadA*7.10 + Y123H_Y147T_Q154R_I76Y)
ABE8.14-d	TadA*8.14	Heterodimer_(WT) + (TadA*7.10 + I76Y_V82S)
ABE8.15-d	TadA*8.15	Heterodimer_(WT) + (TadA*7.10 + V82S_ Y147R)
ABE8.16-d	TadA*8.16	Heterodimer_(WT) + (TadA*7.10 + V82S_Y123H_Y147R)
ABE8.17-d	TadA*8.17	Heterodimer_(WT) + (TadA*7.10 + V82S_Q154R)
ABE8.18-d	TadA*8.18	Heterodimer_(WT) + (TadA*7.10 + V82S_Y123H_Q154R)
ABE8.19-d	TadA*8.19	Heterodimer_(WT) + (TadA*7.10 + V82S_Y123H_Y147R_Q154R)
ABE8.20-d	TadA*8.20	Heterodimer_(WT) + (TadA*7.10 + I76Y_V82S_Y123H_Y147R_Q154R)
ABE8.21-d	TadA*8.21	Heterodimer_(WT) + (TadA*7.10 + Y147R_Q154S)
ABE8.22-d	TadA*8.22	Heterodimer_(WT) + (TadA*7.10 + V82S_Q154S)
ABE8.23-d	TadA*8.23	Heterodimer_(WT) + (TadA*7.10 + V82S_Y123H)
ABE8.24-d	TadA*8.24	Heterodimer_(WT) + (TadA*7.10 + V82S_Y123H_Y147T)

In some embodiments, the ABE8 is ABE8a-m, which has a monomeric construct containing TadA*7.10 with R26C, A109S, T111R, D119N, H122N, Y147D, F149Y, T166I, and D167N mutations (TadA*8a). In some embodiments, the ABE8 is ABE8b-m, which has a monomeric construct containing TadA*7.10 with V88A, A109S, T111R, D119N, H122N, F149Y, T166I, and D167N mutations (TadA*8b). In some embodiments, the ABE8 is ABE8c-m, which has a monomeric construct containing TadA*7.10 with R26C, A109S, T111R, D119N, H122N, F149Y, T166I, and D167N mutations (TadA*8c). In some embodiments, the ABE8 is ABE8d-m, which has a monomeric construct containing TadA*7.10 with V88A, T111R, D119N, and F149Y mutations (TadA*8d). In some embodiments, the ABE8 is ABE8e-m, which has a monomeric construct containing

TadA*7.10 with A109S, T111R, D119N, H122N, Y147D, F149Y, T166I, and D167N mutations (TadA*8e).

In some embodiments, the ABE8 is ABE8a-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with R26C, A109S, T111R, D119, H122N, Y147D, F149Y, T166I, and D167N mutations (TadA*8a). In some embodiments, the ABE8 is ABE8b-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V88A, A109S, T111R, D119N, H122N, F149Y, T166I, and D167N mutations (TadA*8b). In some embodiments, the ABE8 is ABE8c-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with R26C, A109S, T111R, D119N, H122N, F149Y, T166I, and D167N mutations (TadA*8c). In some embodiments, the ABE8 is ABE8d-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V88A, T111R, D119N, and F149Y mutations (TadA*8d). In some embodiments, the ABE8 is ABE8e-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with A109S, T111R, D119N, H122N, Y147D, F149Y, T166I, and D167N mutations (TadA*8e).

In some embodiments, the ABE8 is ABE8a-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with R26C, A109S, T111R, D119, H122N, Y147D, F149Y, T166I, and D167N mutations (TadA*8a). In some embodiments, the ABE8 is ABE8b-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V88A, A109S, T111R, D119N, H122N, F149Y, T166I, and D167N mutations (TadA*8b). In some embodiments, the ABE8 is ABE8c-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with R26C, A109S, T111R, D119N, H122N, F149Y, T166I, and D167N mutations (TadA*8c). In some embodiments, the ABE8 is ABE8d-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V88A, T111R, D119N, and F149Y mutations (TadA*8d). In some embodiments, the ABE8 is ABE8e-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with A109S, T111R, D119N, H122N, Y147D, F149Y, T166I, and D167N mutations (TadA*8e).

In some embodiments, the ABE is ABE8a-m, ABE8b-m, ABE8c-m, ABE8d-m, ABE8e-m, ABE8a-d, ABE8b-d, ABE8c-d, ABE8d-d, or ABE8e-d, as shown in Table 12 below. In some embodiments, the ABE is ABE8e-m or ABE8e-d. ABE8e shows efficient adenine base editing activity and low indel formation when used with Cas homologues other than SpCas9, for example, SaCas9, SaCas9-KKH, Cas12a homologues, *e.g.*, LbCas12a, enAs-Cas12a, SpCas9-NG and circularly permuted CP1028-SpCas9 and CP1041-SpCas9. In

addition to the mutations shown for ABE8e in Table 12, off-target RNA and DNA editing were reduced by introducing a V106W substitution into the TadA domain (as described in M. Richter *et al.*, 2020, Nature Biotechnology, doi.org/10.1038/s41587-020-0453-z, the entire contents of which are incorporated by reference herein).

Table 12: Additional Adenosine Base Editor 8 Variants. In the table, "monomer" indicates an ABE comprising a single TadA*7.10 comprising the indicated alterations and "heterodimer" indicates an ABE comprising a TadA*7.10 comprising the indicated alterations fused to an *E. coli* TadA adenosine deaminase.

ABE8 Base Editor	Adenosine Deaminase	Adenosine Deaminase Description
ABE8a-m	TadA*8a	Monomer_TadA*7.10 + R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N
ABE8b-m	TadA*8b	Monomer_TadA*7.10 + V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N
ABE8c-m	TadA*8c	Monomer_TadA*7.10 + R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N
ABE8d-m	TadA*8d	Monomer_TadA*7.10 + V88A + T111R + D119N + F149Y
ABE8e-m	TadA*8e	Monomer_TadA*7.10 + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N
ABE8a-d	TadA*8a	Heterodimer_(WT) + (TadA*7.10 + R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N)
ABE8b-d	TadA*8b	Heterodimer_(WT) + (TadA*7.10 + V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N)
ABE8c-d	TadA*8c	Heterodimer_(WT) + (TadA*7.10 + R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N)
ABE8d-d	TadA*8d	Heterodimer_(WT) + (TadA*7.10 + V88A + T111R + D119N + F149Y)
ABE8e-d	TadA*8e	Heterodimer_(WT) + (TadA*7.10 + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N)

In some embodiments, base editors (*e.g.*, ABE8) are generated by cloning an adenosine deaminase variant (*e.g.*, TadA*8) into a scaffold that includes a circular permutant Cas9 (*e.g.*, CP5 or CP6) and a bipartite nuclear localization sequence. In some embodiments, the base editor (*e.g.*, ABE7.9, ABE7.10, or ABE8) is an NGC PAM CP5 variant (*S. pyogenes* Cas9 or spVRQR Cas9). In some embodiments, the base editor (*e.g.*, ABE7.9, ABE7.10, or ABE8) is an AGA PAM CP5 variant (*S. pyogenes* Cas9 or spVRQR Cas9). In some embodiments, the base editor (*e.g.*, ABE7.9, ABE7.10, or ABE8) is an NGC PAM CP6 variant (*S. pyogenes* Cas9 or spVRQR Cas9). In some embodiments, the base editor (*e.g.*

ABE7.9, ABE7.10, or ABE8) is an AGA PAM CP6 variant (*S. pyogenes* Cas9 or spVRQR Cas9).

In some embodiments, the ABE has a genotype as shown in Table 13 below.

Table 13. Genotypes of ABEs

	23	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	147	152	155	156	157	161
ABE7.9	L	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	P	V	F	N	K
ABE7.10	R	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K

As shown in Table 14 below, genotypes of 40 ABE8s are described. Residue positions in the evolved *E. coli* TadA portion of ABE are indicated. Mutational changes in ABE8 are shown when distinct from ABE7.10 mutations. In some embodiments, the ABE has a genotype of one of the ABEs as shown in Table 14 below.

Table 14. Residue Identity in Evolved TadA

	23	36	48	51	76	82	84	106	108	123	146	147	152	154	155	156	157	166
ABE7.10	R	L	A	L	I	V	F	V	N	Y	C	Y	P	Q	V	F	N	T
ABE8.1-m												T						
ABE8.2-m												R						
ABE8.3-m														S				
ABE8.4-m										H								
ABE8.5-m						S												
ABE8.6-m																		R
ABE8.7-m														R				
ABE8.8-m										H		R		R				
ABE8.9-m					Y							R		R				
ABE8.10-m												R		R				R
ABE8.11-m												T		R				
ABE8.12-m												T		S				
ABE8.13-m					Y					H		R		R				
ABE8.14-m					Y	S												
ABE8.15-m						S						R						
ABE8.16-m						S				H		R						
ABE8.17-m						S								R				
ABE8.18-m						S				H				R				
ABE8.19-m						S				H		R		R				
ABE8.20-m					Y	S				H		R		R				
ABE8.21-m												R		S				
ABE8.22-m						S								S				

ABE8.23-m						S				H								
ABE8.24-m						S				H		T						
ABE8.1-d												T						
ABE8.2-d												R						
ABE8.3-d														S				
ABE8.4-d										H								
ABE8.5-d						S												
ABE8.6-d																		R
ABE8.7-d														R				
ABE8.8-d										H		R		R				
ABE8.9-d					Y							R		R				
ABE8.10-d												R		R				R
ABE8.11-d												T		R				
ABE8.12-d												T		S				
ABE8.13-d					Y					H		R		R				
ABE8.14-d					Y	S												
ABE8.15-d						S						R						
ABE8.16-d						S				H		R						
ABE8.17-d						S								R				
ABE8.18-d						S				H				R				
ABE8.19-d						S				H		R		R				
ABE8.20-d					Y	S				H		R		R				
ABE8.21-d												R		S				
ABE8.22-d						S								S				
ABE8.23-d						S				H								
ABE8.24-d						S				H		T						

In some embodiments, the base editor is ABE8.1, which comprises or consists essentially of the following sequence or a fragment thereof having adenosine deaminase activity:

ABE8.1_Y147T_CP5_NGC PAM_monomer

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD***SGGSSGGSSGSETPGTSES***
ATPESSGGSSGGSEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK
GRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPT
VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH
KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAF

KYFDTTIARKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD*GGSGGSGGSGGSGGSGGS*
GGMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKL
FIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSL
GLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRV
NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQE
EFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPF
LKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIER
MTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTN
RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILED
IVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK
VVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGK
SDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIT
KHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN
AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ*EGADKRTADGSEFESPKKKRKV*
(SEQ ID NO: 419).

In the above sequence, the plain text denotes an adenosine deaminase sequence, bold sequence indicates sequence derived from Cas9, the italicized sequence denotes a linker sequence, and the underlined sequence denotes a bipartite nuclear localization sequence. Other ABE8 sequences are provided in the attached sequence listing (SEQ ID NOs: 420-442).

In some embodiments, the base editor is a ninth generation ABE (ABE9). In some embodiments, the ABE9 contains a TadA*9 variant. ABE9 base editors include an adenosine deaminase variant comprising an amino acid sequence, which contains alterations relative to an ABE 7*10 reference sequence, as described herein. Exemplary ABE9 variants are listed in Table 15. Details of ABE9 base editors are described in International PCT Application No. PCT/2020/049975, which is incorporated herein by reference for its entirety.

Table 15. Adenosine Base Editor 9 (ABE9) Variants. In the table, "monomer" indicates an ABE comprising a single TadA*7.10 comprising the indicated alterations

and "heterodimer" indicates an ABE comprising a TadA*7.10 comprising the indicated alterations fused to an *E. coli* TadA adenosine deaminase.

ABE9 Description	Alterations
ABE9.1_monomer	E25F, V82S, Y123H, T133K, Y147R, Q154R
ABE9.2_monomer	E25F, V82S, Y123H, Y147R, Q154R
ABE9.3_monomer	V82S, Y123H, P124W, Y147R, Q154R
ABE9.4_monomer	L51W, V82S, Y123H, C146R, Y147R, Q154R
ABE9.5_monomer	P54C, V82S, Y123H, Y147R, Q154R
ABE9.6_monomer	Y73S, V82S, Y123H, Y147R, Q154R
ABE9.7_monomer	N38G, V82T, Y123H, Y147R, Q154R
ABE9.8_monomer	R23H, V82S, Y123H, Y147R, Q154R
ABE9.9_monomer	R21N, V82S, Y123H, Y147R, Q154R
ABE9.10_monomer	V82S, Y123H, Y147R, Q154R, A158K
ABE9.11_monomer	N72K, V82S, Y123H, D139L, Y147R, Q154R,
ABE9.12_monomer	E25F, V82S, Y123H, D139M, Y147R, Q154R
ABE9.13_monomer	M70V, V82S, M94V, Y123H, Y147R, Q154R
ABE9.14_monomer	Q71M, V82S, Y123H, Y147R, Q154R
ABE9.15_heterodimer	E25F, V82S, Y123H, T133K, Y147R, Q154R
ABE9.16_heterodimer	E25F, V82S, Y123H, Y147R, Q154R
ABE9.17_heterodimer	V82S, Y123H, P124W, Y147R, Q154R
ABE9.18_heterodimer	L51W, V82S, Y123H, C146R, Y147R, Q154R
ABE9.19_heterodimer	P54C, V82S, Y123H, Y147R, Q154R
ABE9.2_heterodimer	Y73S, V82S, Y123H, Y147R, Q154R
ABE9.21_heterodimer	N38G, V82T, Y123H, Y147R, Q154R
ABE9.22_heterodimer	R23H, V82S, Y123H, Y147R, Q154R
ABE9.23_heterodimer	R21N, V82S, Y123H, Y147R, Q154R
ABE9.24_heterodimer	V82S, Y123H, Y147R, Q154R, A158K
ABE9.25_heterodimer	N72K, V82S, Y123H, D139L, Y147R, Q154R,
ABE9.26_heterodimer	E25F, V82S, Y123H, D139M, Y147R, Q154R
ABE9.27_heterodimer	M70V, V82S, M94V, Y123H, Y147R, Q154R
ABE9.28_heterodimer	Q71M, V82S, Y123H, Y147R, Q154R
ABE9.29_monomer	E25F_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.30_monomer	I76Y_V82T_Y123H_Y147R_Q154R
ABE9.31_monomer	N38G_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.32_monomer	N38G_I76Y_V82T_Y123H_Y147R_Q154R
ABE9.33_monomer	R23H_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.34_monomer	P54C_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.35_monomer	R21N_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.36_monomer	I76Y_V82S_Y123H_D138M_Y147R_Q154R
ABE9.37_monomer	Y72S_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.38_heterodimer	E25F_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.39_heterodimer	I76Y_V82T_Y123H_Y147R_Q154R
ABE9.40_heterodimer	N38G_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.41_heterodimer	N38G_I76Y_V82T_Y123H_Y147R_Q154R
ABE9.42_heterodimer	R23H_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.43_heterodimer	P54C_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.44_heterodimer	R21N_I76Y_V82S_Y123H_Y147R_Q154R

ABE9.45_heterodimer	I76Y_V82S_Y123H_D138M_Y147R_Q154R
ABE9.46_heterodimer	Y72S_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.47_monomer	N72K_V82S, Y123H, Y147R, Q154R
ABE9.48_monomer	Q71M_V82S, Y123H, Y147R, Q154R
ABE9.49_monomer	M70V,V82S, M94V, Y123H, Y147R, Q154R
ABE9.50_monomer	V82S, Y123H, T133K, Y147R, Q154R
ABE9.51_monomer	V82S, Y123H, T133K, Y147R, Q154R, A158K
ABE9.52_monomer	M70V,Q71M,N72K,V82S, Y123H, Y147R, Q154R
ABE9.53_heterodimer	N72K_V82S, Y123H, Y147R, Q154R
ABE9.54_heterodimer	Q71M_V82S, Y123H, Y147R, Q154R
ABE9.55_heterodimer	M70V,V82S, M94V, Y123H, Y147R, Q154R
ABE9.56_heterodimer	V82S, Y123H, T133K, Y147R, Q154R
ABE9.57_heterodimer	V82S, Y123H, T133K, Y147R, Q154R, A158K
ABE9.58_heterodimer	M70V, Q71M, N72K, V82S, Y123H, Y147R, Q154R

In some embodiments, the base editor includes an adenosine deaminase variant comprising an amino acid sequence, which contains alterations relative to an ABE 7*10 reference sequence, as described herein. The term "monomer" as used in **Table 15.1** refers to a monomeric form of TadA*7.10 comprising the alterations described. The term "heterodimer" as used in **Table 15.1** refers to the specified wild-type *E. coli* TadA adenosine deaminase fused to a TadA*7.10 comprising the alterations as described.

Table 15.1. Adenosine Deaminase Base Editor Variants

ABE	Adenosine Deaminase	Adenosine Deaminase Description
ABE-605m	MSP605	monomer_TadA*7.10 + V82G + Y147T + Q154S
ABE-680m	MSP680	monomer_TadA*7.10 + I76Y + V82G + Y147T + Q154S
ABE-823m	MSP823	monomer_TadA*7.10 + L36H + V82G + Y147T + Q154S + N157K
ABE-824m	MSP824	monomer_TadA*7.10 + V82G + Y147D + F149Y + Q154S + D167N
ABE-825m	MSP825	monomer_TadA*7.10 + L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N
ABE-827m	MSP827	monomer_TadA*7.10 + L36H + I76Y + V82G + Y147T + Q154S + N157K
ABE-828m	MSP828	monomer_TadA*7.10 + I76Y + V82G + Y147D + F149Y + Q154S + D167N
ABE-829m	MSP829	monomer_TadA*7.10 + L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N
ABE-605d	MSP605	heterodimer_(WT)+(TadA*7.10 + V82G + Y147T + Q154S)
ABE-680d	MSP680	heterodimer_(WT)+(TadA*7.10 + I76Y + V82G + Y147T + Q154S)
ABE-823d	MSP823	heterodimer_(WT)+(TadA*7.10 + L36H + V82G + Y147T + Q154S + N157K)

ABE-824d	MSP824	heterodimer_(WT)+(TadA*7.10 + V82G + Y147D + F149Y + Q154S + D167N)
ABE-825d	MSP825	heterodimer_(WT)+(TadA*7.10 + L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N)
ABE-827d	MSP827	heterodimer_(WT)+(TadA*7.10 + L36H + I76Y + V82G + Y147T + Q154S + N157K)
ABE-828d	MSP828	heterodimer_(WT)+(TadA*7.10 + I76Y + V82G + Y147D + F149Y + Q154S + D167N)
ABE-829d	MSP829	heterodimer_(WT)+(TadA*7.10 + L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N)

In some embodiments, the base editor comprises a domain comprising all or a portion of a uracil glycosylase inhibitor (UGI). In some embodiments, the base editor comprises a domain comprising all or a portion of a nucleic acid polymerase. In some embodiments, a base editor can comprise as a domain all or a portion of a nucleic acid polymerase (NAP). For example, a base editor can comprise all or a portion of a eukaryotic NAP. In some embodiments, a NAP or portion thereof incorporated into a base editor is a DNA polymerase. In some embodiments, a NAP or portion thereof incorporated into a base editor has translesion polymerase activity. In some embodiments, a NAP or portion thereof incorporated into a base editor is a translesion DNA polymerase. In some embodiments, a NAP or portion thereof incorporated into a base editor is a Rev7, Rev1 complex, polymerase iota, polymerase kappa, or polymerase eta. In some embodiments, a NAP or portion thereof incorporated into a base editor is a eukaryotic polymerase alpha, beta, gamma, delta, epsilon, gamma, eta, iota, kappa, lambda, mu, or nu component. In some embodiments, a NAP or portion thereof incorporated into a base editor comprises an amino acid sequence that is at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 99.5% identical to a nucleic acid polymerase (*e.g.*, a translesion DNA polymerase). In some embodiments, a nucleic acid polymerase or portion thereof incorporated into a base editor is a translesion DNA polymerase.

In some embodiments, a domain of the base editor can comprise multiple domains. For example, the base editor comprising a polynucleotide programmable nucleotide binding domain derived from Cas9 can comprise a REC lobe and an NUC lobe corresponding to the REC lobe and NUC lobe of a wild-type or natural Cas9. In another example, the base editor can comprise one or more of a RuvCI domain, BH domain, REC1 domain, REC2 domain, RuvCII domain, L1 domain, HNH domain, L2 domain, RuvCIII domain, WED domain, TOPO domain or CTD domain. In some embodiments, one or more domains of the base editor comprise a mutation (*e.g.*, substitution, insertion, deletion) relative to a wild-type

version of a polypeptide comprising the domain. For example, an HNH domain of a polynucleotide programmable DNA binding domain can comprise an H840A substitution. In another example, a RuvCI domain of a polynucleotide programmable DNA binding domain can comprise a D10A substitution.

Different domains (*e.g.*, adjacent domains) of the base editor disclosed herein can be connected to each other with or without the use of one or more linker domains (*e.g.*, an XTEN linker domain). In some embodiments, a linker domain can be a bond (*e.g.*, covalent bond), chemical group, or a molecule linking two molecules or moieties, *e.g.*, two domains of a fusion protein, such as, for example, a first domain (*e.g.*, Cas9-derived domain) and a second domain (*e.g.*, an adenosine deaminase domain or a cytidine deaminase domain). In some embodiments, a linker is a covalent bond (*e.g.*, a carbon-carbon bond, disulfide bond, carbon-hetero atom bond, etc.). In certain embodiments, a linker is a carbon nitrogen bond of an amide linkage. In certain embodiments, a linker is a cyclic or acyclic, substituted or unsubstituted, branched or unbranched aliphatic or heteroaliphatic linker. In certain embodiments, a linker is polymeric (*e.g.*, polyethylene, polyethylene glycol, polyamide, polyester, etc.). In certain embodiments, a linker comprises a monomer, dimer, or polymer of aminoalkanoic acid. In some embodiments, a linker comprises an aminoalkanoic acid (*e.g.*, glycine, ethanoic acid, alanine, beta-alanine, 3-aminopropanoic acid, 4-aminobutanoic acid, 5-pentanoic acid, etc.). In some embodiments, a linker comprises a monomer, dimer, or polymer of aminohexanoic acid (Ahx). In certain embodiments, a linker is based on a carbocyclic moiety (*e.g.*, cyclopentane, cyclohexane). In other embodiments, a linker comprises a polyethylene glycol moiety (PEG). In certain embodiments, a linker comprises an aryl or heteroaryl moiety. In certain embodiments, the linker is based on a phenyl ring. A linker can include functionalized moieties to facilitate attachment of a nucleophile (*e.g.*, thiol, amino) from the peptide to the linker. Any electrophile can be used as part of the linker. Exemplary electrophiles include, but are not limited to, activated esters, activated amides, Michael acceptors, alkyl halides, aryl halides, acyl halides, and isothiocyanates. In some embodiments, a linker joins a gRNA binding domain of an RNA-programmable nuclease, including a Cas9 nuclease domain, and the catalytic domain of a nucleic acid editing protein. In some embodiments, a linker joins a dCas9 and a second domain (*e.g.*, UGI, etc.).

Linkers

In certain embodiments, linkers may be used to link any of the peptides or peptide domains of the invention. The linker may be as simple as a covalent bond, or it may be a

polymeric linker many atoms in length. In certain embodiments, the linker is a polypeptide or based on amino acids. In certain embodiments, polypeptide or amino acid-based linkers may be encoded by any of the polynucleotides of the invention. In some embodiments, a polynucleotide encoding a deaminase domain and/or a nucleic acid programmable DNA binding protein (napDNAbp) domain, or a fragment thereof, comprises a linker polynucleotide sequence. In some embodiments, a polynucleotide encoding a deaminase domain and/or a nucleic acid programmable DNA binding protein (napDNAbp) domain, or a fragment thereof, and a linker polynucleotide sequence, includes an intron inserted within an open reading frame. In some embodiments, the intron is inserted within the linker polynucleotide sequence.

In other embodiments, the linker is not peptide-like. In certain embodiments, the linker is a covalent bond (*e.g.*, a carbon-carbon bond, disulfide bond, carbon-heteroatom bond, *etc.*). In certain embodiments, the linker is a carbon-nitrogen bond of an amide linkage. In certain embodiments, the linker is a cyclic or acyclic, substituted or unsubstituted, branched or unbranched aliphatic or heteroaliphatic linker. In certain embodiments, the linker is polymeric (*e.g.*, polyethylene, polyethylene glycol, polyamide, polyester, *etc.*). In certain embodiments, the linker comprises a monomer, dimer, or polymer of aminoalkanoic acid. In certain embodiments, the linker comprises an aminoalkanoic acid (*e.g.*, glycine, ethanoic acid, alanine, beta-alanine, 3-aminopropanoic acid, 4-aminobutanoic acid, 5-pentanoic acid, *etc.*). In certain embodiments, the linker comprises a monomer, dimer, or polymer of aminohexanoic acid (Ahx). In certain embodiments, the linker is based on a carbocyclic moiety (*e.g.*, cyclopentane, cyclohexane). In other embodiments, the linker comprises a polyethylene glycol moiety (PEG). In other embodiments, the linker comprises amino acids. In certain embodiments, the linker comprises a peptide. In certain embodiments, the linker comprises an aryl or heteroaryl moiety. In certain embodiments, the linker is based on a phenyl ring. The linker may include functionalized moieties to facilitate attachment of a nucleophile (*e.g.*, thiol, amino) from the peptide to the linker. Any electrophile may be used as part of the linker. Exemplary electrophiles include, but are not limited to, activated esters, activated amides, Michael acceptors, alkyl halides, aryl halides, acyl halides, and isothiocyanates.

Typically, a linker is positioned between, or flanked by, two groups, molecules, or other moieties and connected to each one via a covalent bond, thus connecting the two. In some embodiments, a linker is an amino acid or a plurality of amino acids (*e.g.*, a peptide or protein). In some embodiments, a linker is an organic molecule, group, polymer, or chemical

moiety. In some embodiments, a linker is 2-100 amino acids in length, for example, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, or 150-200 amino acids in length. In some embodiments, the linker is about 3 to about 104 (*e.g.*, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100) amino acids in length. Longer or shorter linkers are also contemplated.

In some embodiments, any of the fusion proteins provided herein, comprise a cytidine or adenosine deaminase and a Cas9 domain that are fused to each other via a linker. Various linker lengths and flexibilities between the cytidine or adenosine deaminase and the Cas9 domain can be employed (*e.g.*, ranging from very flexible linkers of the form (GGGS)n (SEQ ID NO: 334), (GGGGS)n (SEQ ID NO: 335), and (G)n to more rigid linkers of the form (EAAAK)n (SEQ ID NO: 336), (SGGS)n (SEQ ID NO: 443), SGSETPGTSESATPES (SEQ ID NO: 337) (see, *e.g.*, Guilinger JP, et al. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nat. Biotechnol. 2014; 32(6): 577-82; the entire contents are incorporated herein by reference) and (XP)n) in order to achieve the optimal length for activity for the cytidine or adenosine deaminase nucleobase editor. In some embodiments, n is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15. In some embodiments, the linker comprises a (GGS)n motif, wherein n is 1, 3, or 7. In some embodiments, cytidine deaminase or adenosine deaminase and the Cas9 domain of any of the fusion proteins provided herein are fused via a linker comprising the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 237), which can also be referred to as the XTEN linker.

In some embodiments, the domains of the base editor are fused via a linker that comprises the amino acid sequence of:
SGGSSGSETPGTSESATPESSGGS (SEQ ID NO: 444),
SGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID NO: 445), or
GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPS
EGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS (SEQ ID NO: 446).

In some embodiments, domains of the base editor are fused via a linker comprising the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 237), which may also be referred to as the XTEN linker. In some embodiments, a linker comprises the amino acid sequence SGGS. In some embodiments, the linker is 24 amino acids in length. In some

embodiments, the linker comprises the amino acid sequence SGGSSGGSSGSETPGTSESATPES (SEQ ID NO: 447). In some embodiments, the linker is 40 amino acids in length. In some embodiments, the linker comprises the amino acid sequence: SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGS (SEQ ID NO: 448). In some embodiments, the linker is 64 amino acids in length. In some embodiments, the linker comprises the amino acid sequence: SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID NO: 449). In some embodiments, the linker is 92 amino acids in length. In some embodiments, the linker comprises the amino acid sequence: PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATS (SEQ ID NO: 450).

In some embodiments, a linker comprises a plurality of proline residues and is 5-21, 5-14, 5-9, 5-7 amino acids in length, *e.g.*, PAPAP (SEQ ID NO: 451), PAPAPA (SEQ ID NO: 452), PAPAPAP (SEQ ID NO: 453), PAPAPAPA (SEQ ID NO: 454), P(AP)4 (SEQ ID NO: 455), P(AP)7 (SEQ ID NO: 456), P(AP)10 (SEQ ID NO: 457) (see, *e.g.*, Tan J, Zhang F, Karcher D, Bock R. Engineering of high-precision base editors for site-specific single nucleotide replacement. Nat Commun. 2019 Jan 25;10(1):439; the entire contents are incorporated herein by reference). Such proline-rich linkers are also termed "rigid" linkers.

In another embodiment, the base editor system comprises a component (protein) that interacts non-covalently with a deaminase (DNA deaminase), *e.g.*, an adenosine or a cytidine deaminase, and transiently attracts the adenosine or cytidine deaminase to the target nucleobase in a target polynucleotide sequence for specific editing, with minimal or reduced bystander or target-adjacent effects. Such a non-covalent system and method involving deaminase-interacting proteins serves to attract a DNA deaminase to a particular genomic target nucleobase and decouples the events of on-target and target-adjacent editing, thus enhancing the achievement of more precise single base substitution mutations. In an embodiment, the deaminase-interacting protein binds to the deaminase (*e.g.*, adenosine deaminase or cytidine deaminase) without blocking or interfering with the active (catalytic) site of the deaminase from engaging the target nucleobase (*e.g.*, adenosine or cytidine, respectively). Such as system, termed "MagnEdit," involves interacting proteins tethered to a Cas9 and gRNA complex and can attract a co-expressed adenosine or cytidine deaminase (either exogenous or endogenous) to edit a specific genomic target site, and is described in McCann, J. et al., 2020, "MagnEdit – interacting factors that recruit DNA-editing enzymes to

single base targets," Life-Science-Alliance, Vol. 3, No. 4 (e201900606), (doi 10.26508/lsa.201900606), the contents of which are incorporated by reference herein in their entirety. In an embodiment, the DNA deaminase is an adenosine deaminase variant (*e.g.*, TadA*8) as described herein.

In another embodiment, a system called "Suntag," involves non-covalently interacting components used for recruiting protein (*e.g.*, adenosine deaminase or cytidine deaminase) components, or multiple copies thereof, of base editors to polynucleotide target sites to achieve base editing at the site with reduced adjacent target editing, for example, as described in Tanenbaum, M.E. et al., "A protein tagging system for signal amplification in gene expression and fluorescence imaging," Cell. 2014 October 23; 159(3): 635–646. doi:10.1016/j.cell.2014.09.039; and in Huang, Y.-H. et al., 2017, "DNA epigenome editing using CRISPR-Cas SunTag-directed DNMT3A," Genome Biol 18: 176. doi:10.1186/s13059-017-1306-z, the contents of each of which are incorporated by reference herein in their entirety. In an embodiment, the DNA deaminase is an adenosine deaminase variant (*e.g.*, TadA*8) as described herein.

Nucleic Acid Programmable DNA Binding Proteins with Guide RNAs

Provided herein are compositions and methods for base editing and/or inactivating a base editor in cells. Further provided herein are compositions comprising a guide polynucleic acid sequence, *e.g.* a guide RNA sequence, or a combination of 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more guide RNAs as provided herein. In some embodiments, a composition for base editing as provided herein further comprises a polynucleotide that encodes a base editor, *e.g.* a C-base editor or an A-base editor. For example, a composition for base editing may comprise a mRNA sequence encoding a BE, a BE4, an ABE, and a combination of one or more guide RNAs as provided. In some embodiments, the polynucleotide that encodes a base editor includes a heterologous intron. A composition for base editing may comprise a base editor polypeptide and a combination of one or more of any guide RNAs provided herein. Such a composition may be used to effect base editing or to inactivate a base editor in a cell through different delivery approaches, for example, electroporation, nucleofection, viral transduction or transfection. In some embodiments, the composition for base editing and/or inactivating a base editor comprises an mRNA sequence that encodes a base editor and a combination of one or more guide RNA sequences provided herein for electroporation. In some embodiments, the mRNA sequence that encodes a base editor includes a heterologous intron.

Some aspects of this disclosure provide complexes comprising any of the fusion proteins provided herein, and a guide RNA bound to a nucleic acid programmable DNA binding protein (napDNAbp) domain (*e.g.,* a Cas9 (*e.g.*, a dCas9, a nuclease active Cas9, or a Cas9 nickase) or Cas12) of the fusion protein. These complexes are also termed ribonucleoproteins (RNPs). In some embodiments, the guide nucleic acid (*e.g.*, guide RNA) is from 15-100 nucleotides long and comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence. In some embodiments, the guide RNA is 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 nucleotides long. In some embodiments, the guide RNA comprises a sequence of 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 contiguous nucleotides that is complementary to a target sequence. In some embodiments, the target sequence is a DNA sequence. In some embodiments, the target sequence is an RNA sequence. In some embodiments, the target sequence is a sequence in the genome of a bacteria, yeast, fungi, insect, plant, or animal. In some embodiments, the target sequence is a sequence in the genome of a human. In some embodiments, the 3' end of the target sequence is immediately adjacent to a canonical PAM sequence (NGG). In some embodiments, the 3' end of the target sequence is immediately adjacent to a non-canonical PAM sequence (*e.g.*, a sequence listed in Table 6 or 5'-NAA-3'). In some embodiments, the guide nucleic acid (*e.g.*, guide RNA) is complementary to a sequence in a gene of interest (*e.g.*, a gene associated with a disease or disorder).

Some aspects of this disclosure provide methods of using the fusion proteins, or complexes provided herein. For example, some aspects of this disclosure provide methods comprising contacting a DNA molecule with any of the fusion proteins provided herein, and with at least one guide RNA, wherein the guide RNA is about 15-100 nucleotides long and comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence. In some embodiments, the 3' end of the target sequence is immediately adjacent to an AGC, GAG, TTT, GTG, or CAA sequence. In some embodiments, the 3' end of the target sequence is immediately adjacent to an NGA, NGCG, NGN, NNGRRT, NNNRRT, NGCG, NGCN, NGTN, NGTN, NGTN, or 5' (TTTV) sequence. In some embodiments, the 3' end of the target sequence is immediately adjacent to an *e.g.*, TTN, DTTN, GTTN, ATTN, ATTC, DTTNT, WTTN, HATY, TTTN, TTTV, TTTC, TG, RTR, or YTN PAM site.

It will be understood that the numbering of the specific positions or residues in the respective sequences depends on the particular protein and numbering scheme used.

Numbering might differ, *e.g.*, in precursors of a mature protein and the mature protein itself, and differences in sequences from species to species may affect numbering. One of skill in the art will be able to identify the respective residue in any homologous protein and in the respective encoding nucleic acid by methods well known in the art, *e.g.*, by sequence alignment and determination of homologous residues.

It will be apparent to those of skill in the art that in order to target any of the fusion proteins disclosed herein, to a target site, *e.g.*, a site comprising a mutation to be edited, it is typically necessary to co-express the fusion protein together with a guide RNA. As explained in more detail elsewhere herein, a guide RNA typically comprises a tracrRNA framework allowing for napDNAbp (*e.g.*, Cas9 or Cas12) binding, and a guide sequence, which confers sequence specificity to the napDNAbp:nucleic acid editing enzyme/domain fusion protein. Alternatively, the guide RNA and tracrRNA may be provided separately, as two nucleic acid molecules. In some embodiments, the guide RNA comprises a structure, wherein the guide sequence comprises a sequence that is complementary to the target sequence. The guide sequence is typically 20 nucleotides long. The sequences of suitable guide RNAs for targeting napDNAbp:nucleic acid editing enzyme/domain fusion proteins to specific genomic target sites will be apparent to those of skill in the art based on the instant disclosure. Such suitable guide RNA sequences typically comprise guide sequences that are complementary to a nucleic sequence within 50 nucleotides upstream or downstream of the target nucleotide to be edited. Some exemplary guide RNA sequences suitable for targeting any of the provided fusion proteins to specific target sequences are provided herein.

Distinct portions of sgRNA are predicted to form various features that interact with Cas9 (*e.g.*, SpyCas9) and/or the DNA target. Six conserved modules have been identified within native crRNA:tracrRNA duplexes and single guide RNAs (sgRNAs) that direct Cas9 endonuclease activity (see Briner et al., Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality Mol Cell. 2014 Oct 23;56(2):333-339). The six modules include the spacer responsible for DNA targeting, the upper stem, bulge, lower stem formed by the CRISPR repeat:tracrRNA duplex, the nexus, and hairpins from the 3' end of the tracrRNA. The upper and lower stems interact with Cas9 mainly through sequence-independent interactions with the phosphate backbone. In some embodiments, the upper stem is dispensable. In some embodiments, the conserved uracil nucleotide sequence at the base of the lower stem is dispensable. The bulge participates in specific side-chain interactions with the Rec1 domain of Cas9. The nucleobase of U44 interacts with the side chains of Tyr 325

and His 328, while G43 interacts with Tyr 329. The nexus forms the core of the sgRNA:Cas9 interactions and lies at the intersection between the sgRNA and both Cas9 and the target DNA. The nucleobases of A51 and A52 interact with the side chain of Phe 1105; U56 interacts with Arg 457 and Asn 459; the nucleobase of U59 inserts into a hydrophobic pocket defined by side chains of Arg 74, Asn 77, Pro 475, Leu 455, Phe 446, and Ile 448; C60 interacts with Leu 455, Ala 456, and Asn 459, and C61 interacts with the side chain of Arg 70, which in turn interacts with C15. In some embodiments, one or more of these mutations are made in the bulge and/or the nexus of a sgRNA for a Cas9 (*e.g.*, spyCas9) to optimize sgRNA:Cas9 interactions.

Moreover, the tracrRNA nexus and hairpins are critical for Cas9 pairing and can be swapped to cross orthogonality barriers separating disparate Cas9 proteins, which is instrumental for further harnessing of orthogonal Cas9 proteins. In some embodiments, the nexus and hairpins are swapped to target orthogonal Cas9 proteins. In some embodiments, a sgRNA is dispensed of the upper stem, hairpin 1, and/or the sequence flexibility of the lower stem to design a guide RNA that is more compact and conformationally stable. In some embodiments, the modules are modified to optimize multiplex editing using a single Cas9 with various chimeric guides or by concurrently using orthogonal systems with different combinations of chimeric sgRNAs. Details regarding guide functional modules and methods thereof are described, for example, in Briner et al., Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality Mol Cell. 2014 Oct 23;56(2):333-339, the contents of which is incorporated by reference herein in its entirety.

The domains of the base editor disclosed herein can be arranged in any order. Non-limiting examples of a base editor comprising a fusion protein comprising *e.g.*, a polynucleotide-programmable nucleotide-binding domain (*e.g.*, Cas9 or Cas12) and a deaminase domain (*e.g.*, cytidine or adenosine deaminase) can be arranged as follows:

NH2-[nucleobase editing domain]-Linker1-[nucleobase editing domain]-COOH;
NH2-[deaminase]-Linker1-[nucleobase editing domain]-COOH;
NH2-[deaminase]-Linker1-[nucleobase editing domain]-Linker2-[UGI]-COOH;
NH2-[deaminase]-Linker1-[nucleobase editing domain]-COOH;
NH2-[adenosine deaminase]-Linker1-[nucleobase editing domain]-COOH;
NH2-[nucleobase editing domain]-[deaminase]-COOH;
NH2-[deaminase]-[nucleobase editing domain]-[inosine BER inhibitor]-COOH;
NH2-[deaminase]-[inosine BER inhibitor]-[nucleobase editing domain]-COOH;
NH2-[inosine BER inhibitor]-[deaminase]-[nucleobase editing domain]-COOH;

NH2-[nucleobase editing domain]-[deaminase]-[inosine BER inhibitor]-COOH;

NH2-[nucleobase editing domain]-[inosine BER inhibitor]-[deaminase]-COOH;

NH2-[inosine BER inhibitor]-[nucleobase editing domain]-[deaminase]-COOH;

NH2-[nucleobase editing domain]-Linker1-[deaminase]-Linker2-[nucleobase editing domain]-COOH;

NH2-[nucleobase editing domain]-Linker1-[deaminase]-[nucleobase editing domain]-COOH;

NH2-[nucleobase editing domain]-[deaminase]-Linker2-[nucleobase editing domain]-COOH;

NH2-[nucleobase editing domain]-[deaminase]-[nucleobase editing domain]-COOH;

NH2-[nucleobase editing domain]-Linker1-[deaminase]-Linker2-[nucleobase editing domain]-[inosine BER inhibitor]-COOH;

NH2-[nucleobase editing domain]-Linker1-[deaminase]-[nucleobase editing domain]-[inosine BER inhibitor]-COOH;

NH2-[nucleobase editing domain]-[deaminase]-Linker2-[nucleobase editing domain]-[inosine BER inhibitor]-COOH;

NH2-[nucleobase editing domain]-[deaminase]-[nucleobase editing domain]-[inosine BER inhibitor]-COOH;

NH2-[inosine BER inhibitor]-[nucleobase editing domain]-Linker1-[deaminase]-Linker2-[nucleobase editing domain]-COOH;

NH2-[inosine BER inhibitor]-[nucleobase editing domain]-Linker1-[deaminase]-[nucleobase editing domain]-COOH;

NH2-[inosine BER inhibitor]-[nucleobase editing domain]-[deaminase]-Linker2-[nucleobase editing domain]-COOH; or

NH2-[inosine BER inhibitor]NH2-[nucleobase editing domain]-[deaminase]-[nucleobase editing domain]-COOH.

In some embodiments, the base editing fusion proteins provided herein need to be positioned at a precise location, for example, where a target base is placed within a defined region (*e.g.*, a "deamination window"). In some embodiments, a target can be within a 4-base region. In some embodiments, such a defined target region can be approximately 15 bases upstream of the PAM. See Komor, A.C., et al., "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., et al., "Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); and Komor, A.C., et al., "Improved base

excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017), the entire contents of which are hereby incorporated by reference.

A defined target region can be a deamination window. A deamination window can be the defined region in which a base editor acts upon and deaminates a target nucleotide. In some embodiments, the deamination window is within a 2, 3, 4, 5, 6, 7, 8, 9, or 10 base regions. In some embodiments, the deamination window is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 bases upstream of the PAM.

The base editors of the present disclosure can comprise any domain, feature or amino acid sequence which facilitates the editing of a target polynucleotide sequence. For example, in some embodiments, the base editor comprises a nuclear localization sequence (NLS). In some embodiments, an NLS of the base editor is localized between a deaminase domain and a napDNAbp domain. In some embodiments, an NLS of the base editor is localized C-terminal to a napDNAbp domain.

Non-limiting examples of protein domains which can be included in the fusion protein include a deaminase domain (*e.g.*, adenosine deaminase or cytidine deaminase), a uracil glycosylase inhibitor (UGI) domain, epitope tags, reporter gene sequences, and/or protein domains having one or more of the activities described herein.

A domain may be detected or labeled with an epitope tag, a reporter protein, other binding domains. Non-limiting examples of epitope tags include histidine (His) tags, V5 tags, FLAG tags, influenza hemagglutinin (HA) tags, Myc tags, VSV-G tags, and thioredoxin (Trx) tags. Examples of reporter genes include, but are not limited to, glutathione-5-transferase (GST), horseradish peroxidase (HRP), chloramphenicol acetyltransferase (CAT) beta-galactosidase, beta-glucuronidase, luciferase, green fluorescent protein (GFP), HcRed, DsRed, cyan fluorescent protein (CFP), yellow fluorescent protein (YFP), and autofluorescent proteins including blue fluorescent protein (BFP). Additional protein sequences can include amino acid sequences that bind DNA molecules or bind other cellular molecules, including but not limited to maltose binding protein (MBP), S-tag, Lex A DNA binding domain (DBD) fusions, GAL4 DNA binding domain fusions, and herpes simplex virus (HSV) BP16 protein fusions.

Methods of Using Fusion Proteins Comprising a Cytidine or Adenosine Deaminase and a Cas9 Domain

Some aspects of this disclosure provide methods of using the fusion proteins, or complexes provided herein. For example, some aspects of this disclosure provide methods comprising contacting a DNA molecule with any of the fusion proteins provided herein, and with at least one guide RNA described herein.

In some embodiments, a fusion protein of the invention is used for editing a target gene or polynucleotide sequence of interest. In particular, a cytidine deaminase or adenosine deaminase nucleobase editor described herein is capable of making multiple mutations within a target sequence. These mutations may affect the function of the target. For example, when a cytidine deaminase or adenosine deaminase nucleobase editor is used to target a regulatory region the function of the regulatory region is altered and the expression of the downstream protein is reduced or eliminated. In another example, when a cytidine deaminase or adenosine deaminase nucleobase editor is used to target the splice acceptor or splice donor site in a heterologous intron incorporated into a polynucleotide sequence encoding a base editor, the splicing of the intron is altered and the expression or activity of the base editor is reduced or eliminated.

It will be understood that the numbering of the specific positions or residues in the respective sequences depends on the particular protein and numbering scheme used. Numbering might be different, *e.g.*, in precursors of a mature protein and the mature protein itself, and differences in sequences from species to species may affect numbering. One of skill in the art will be able to identify the respective residue in any homologous protein and in the respective encoding nucleic acid by methods well known in the art, *e.g.*, by sequence alignment and determination of homologous residues.

It will be apparent to those of skill in the art that in order to target any of the fusion proteins comprising a Cas9 domain and a cytidine or adenosine deaminase, as disclosed herein, to a target site, *e.g.*, a site comprising a mutation to be edited, it is typically necessary to co-express the fusion protein together with a guide RNA, *e.g.*, an sgRNA. As explained in more detail elsewhere herein, a guide RNA typically comprises a tracrRNA framework allowing for Cas9 binding, and a guide sequence, which confers sequence specificity to the Cas9:nucleic acid editing enzyme/domain fusion protein. Alternatively, the guide RNA and tracrRNA may be provided separately, as two nucleic acid molecules. In some embodiments, the guide RNA comprises a structure, wherein the guide sequence comprises a sequence that is complementary to the target sequence. The guide sequence is typically 20 nucleotides

long. The sequences of suitable guide RNAs for targeting Cas9:nucleic acid editing enzyme/domain fusion proteins to specific genomic target sites will be apparent to those of skill in the art based on the instant disclosure. Such suitable guide RNA sequences typically comprise guide sequences that are complementary to a nucleic sequence within 50 nucleotides upstream or downstream of the target nucleotide to be edited. Some exemplary guide RNA sequences suitable for targeting any of the provided fusion proteins to specific target sequences are provided herein.

Base Editor Efficiency

In some embodiments, the purpose of the methods provided herein is to alter a gene and/or gene product via gene editing. The nucleobase editing proteins provided herein can be used for gene editing-based human therapeutics *in vitro* or *in vivo*. It will be understood by the skilled artisan that the nucleobase editing proteins provided herein, *e.g.*, the fusion proteins comprising a polynucleotide programmable nucleotide binding domain (*e.g.*, Cas9) and a nucleobase editing domain (*e.g.*, an adenosine deaminase domain or a cytidine deaminase domain) can be used to edit a nucleotide from A to G or C to T. In some embodiments, the base editor is a self-inactivating base editor, where the inactivation is induced by editing an intron present in a polynucleotide encoding the base editor.

Advantageously, base editing systems as provided herein provide genome editing without generating double-strand DNA breaks, without requiring a donor DNA template, and without inducing an excess of stochastic insertions and deletions as CRISPR may do. In some embodiments, the present disclosure provides base editors that efficiently generate an intended mutation, such as a STOP codon, in a nucleic acid (*e.g.*, a nucleic acid within a genome of a subject) without generating a significant number of unintended mutations, such as unintended point mutations. In some embodiments, an intended mutation is a mutation that is generated by a specific base editor (*e.g.*, adenosine base editor or cytidine base editor) bound to a guide polynucleotide (*e.g.*, gRNA), specifically designed to generate the intended mutation. In some embodiments, the intended mutation is in a gene associated with a target antigen associated with a disease or disorder. In some embodiments, the intended mutation is an adenine (A) to guanine (G) point mutation (*e.g.*, SNP) in a gene associated with a target antigen associated with a disease or disorder. In some embodiments, the intended mutation is an adenine (A) to guanine (G) point mutation within the coding region or non-coding region of a gene (*e.g.*, regulatory region or element). In some embodiments, the intended mutation is a cytosine (C) to thymine (T) point mutation (*e.g.*, SNP) in a gene associated with a target

antigen associated with a disease or disorder. In some embodiments, the intended mutation is a cytosine (C) to thymine (T) point mutation within the coding region or non-coding region of a gene (*e.g.*, regulatory region or element). In some embodiments, the intended mutation is a point mutation that generates a STOP codon, for example, a premature STOP codon within the coding region of a gene. In some embodiments, the intended mutation is a mutation that eliminates a stop codon.

In some embodiments, the intended edit is in an intron of a polynucleotide encoding a self-inactivating base editor. In some embodiments, the intended edit is in a splice acceptor or a splice donor site present in the intron of a polynucleotide encoding a self-inactivating base editor. In some embodiments, the intended edit is an adenine (A) to guanine (G) point mutation (*e.g.*, SNP) in an intron of a polynucleotide encoding a self-inactivating base editor. In some embodiments, the intended edit is an adenine (A) to guanine (G) point mutation within the splice acceptor or a splice donor site present in the intron of a polynucleotide encoding a self-inactivating base editor. In some embodiments, the intended edit is a cytosine (C) to thymine (T) point mutation (*e.g.*, SNP) in an intron of a polynucleotide encoding a self-inactivating base editor. In some embodiments, the intended mutation is a cytosine (C) to thymine (T) point mutation within the splice acceptor or a splice donor site present in the intron of a polynucleotide encoding a self-inactivating base editor.

The base editors of the invention advantageously modify a specific nucleotide base encoding a protein without generating a significant proportion of indels. An "indel", as used herein, refers to the insertion or deletion of a nucleotide base within a nucleic acid. Such insertions or deletions can lead to frame shift mutations within a coding region of a gene. In some embodiments, it is desirable to generate base editors that efficiently modify (*e.g.* mutate) a specific nucleotide within a nucleic acid, without generating a large number of insertions or deletions (*i.e.*, indels) in the nucleic acid. In some embodiments, it is desirable to generate base editors that efficiently modify (*e.g.* mutate or methylate) a specific nucleotide within a nucleic acid, without generating a large number of insertions or deletions (*i.e.*, indels) in the nucleic acid. In certain embodiments, any of the base editors provided herein can generate a greater proportion of intended modifications (*e.g.*, methylations) versus indels. In certain embodiments, any of the base editors provided herein can generate a greater proportion of intended modifications (*e.g.*, mutations) versus indels.

In some embodiments, the base editors provided herein are capable of generating a ratio of intended mutations to indels (*i.e.*, intended point mutations:unintended point mutations) that is greater than 1:1. In some embodiments, the base editors provided herein

are capable of generating a ratio of intended mutations to indels that is at least 1.5:1, at least 2:1, at least 2.5:1, at least 3:1, at least 3.5:1, at least 4:1, at least 4.5:1, at least 5:1, at least 5.5:1, at least 6:1, at least 6.5:1, at least 7:1, at least 7.5:1, at least 8:1, at least 10:1, at least 12:1, at least 15:1, at least 20:1, at least 25:1, at least 30:1, at least 40:1, at least 50:1, at least 100:1, at least 200:1, at least 300:1, at least 400:1, at least 500:1, at least 600:1, at least 700:1, at least 800:1, at least 900:1, or at least 1000:1, or more. The number of intended mutations and indels may be determined using any suitable method.

In some embodiments, the base editors provided herein can limit formation of indels in a region of a nucleic acid. In some embodiments, the region is at a nucleotide targeted by a base editor or a region within 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides of a nucleotide targeted by a base editor. In some embodiments, any of the base editors provided herein can limit the formation of indels at a region of a nucleic acid to less than 1%, less than 1.5%, less than 2%, less than 2.5%, less than 3%, less than 3.5%, less than 4%, less than 4.5%, less than 5%, less than 6%, less than 7%, less than 8%, less than 9%, less than 10%, less than 12%, less than 15%, or less than 20%. The number of indels formed at a nucleic acid region may depend on the amount of time a nucleic acid (*e.g.*, a nucleic acid within the genome of a cell) is exposed to a base editor. In some embodiments, a number or proportion of indels is determined after at least 1 hour, at least 2 hours, at least 6 hours, at least 12 hours, at least 24 hours, at least 36 hours, at least 48 hours, at least 3 days, at least 4 days, at least 5 days, at least 7 days, at least 10 days, or at least 14 days of exposing a nucleic acid (*e.g.*, a nucleic acid within the genome of a cell) to a base editor.

Some aspects of the disclosure are based on the recognition that any of the base editors provided herein are capable of efficiently generating an intended mutation in a nucleic acid (*e.g.* a nucleic acid within a genome of a subject) without generating a considerable number of unintended mutations (*e.g.*, spurious off-target editing or bystander editing). In some embodiments, an intended mutation is a mutation that is generated by a specific base editor bound to a gRNA, specifically designed to generate the intended mutation. In some embodiments, the intended mutation is a mutation that generates a stop codon, for example, a premature stop codon within the coding region of a gene. In some embodiments, the intended mutation is a mutation that eliminates a stop codon. In some embodiments, the intended mutation is a mutation that alters the splicing of a gene. In some embodiments, the intended mutation is a mutation that alters the regulatory sequence of a gene (*e.g.*, a gene promotor or gene repressor). In some embodiments, any of the base editors provided herein are capable of generating a ratio of intended mutations to unintended mutations (*e.g.*,

intended mutations:unintended mutations) that is greater than 1:1. In some embodiments, any of the base editors provided herein are capable of generating a ratio of intended mutations to unintended mutations that is at least 1.5:1, at least 2:1, at least 2.5:1, at least 3:1, at least 3.5:1, at least 4:1, at least 4.5:1, at least 5:1, at least 5.5:1, at least 6:1, at least 6.5:1, at least 7:1, at least 7.5:1, at least 8:1, at least 10:1, at least 12:1, at least 15:1, at least 20:1, at least 25:1, at least 30:1, at least 40:1, at least 50:1, at least 100:1, at least 150:1, at least 200:1, at least 250:1, at least 500:1, or at least 1000:1, or more. It should be appreciated that the characteristics of the base editors described herein may be applied to any of the fusion proteins, or methods of using the fusion proteins provided herein.

Base editing is often referred to as a "modification", such as, a genetic modification, a gene modification and modification of the nucleic acid sequence and is clearly understandable based on the context that the modification is a base editing modification. A base editing modification is therefore a modification at the nucleotide base level, for example as a result of the deaminase activity discussed throughout the disclosure, which then results in a change in the gene sequence, and may affect the gene product. In essence therefore, the gene editing modification described herein may result in a modification of the gene, structurally and/or functionally, wherein the expression of the gene product may be modified, for example, the expression of the gene is knocked out; or conversely, enhanced, or, in some circumstances, the gene function or activity may be modified. Using the methods disclosed herein, a base editing efficiency may be determined as the knockdown efficiency of the gene in which the base editing is performed, wherein the base editing is intended to knockdown the expression of the gene. A knockdown level may be validated quantitatively by determining the expression level by any detection assay, such as assay for protein expression level, for example, by flow cytometry; assay for detecting RNA expression such as quantitative RT-PCR, northern blot analysis, or any other suitable assay such as pyrosequencing; and may be validated qualitatively by nucleotide sequencing reactions.

In some embodiments, the modification, *e.g.*, single base edit results in at least 10% reduction of the gene targeted expression. In some embodiments, the base editing efficiency may result in at least 10% reduction of the gene targeted expression. In some embodiments, the base editing efficiency may result in at least 20% reduction of the gene targeted expression. In some embodiments, the base editing efficiency may result in at least 30% reduction of the gene targeted expression. In some embodiments, the base editing efficiency may result in at least 40% reduction of the gene targeted expression. In some embodiments, the base editing efficiency may result in at least 50% reduction of the gene targeted

expression. In some embodiments, the base editing efficiency may result in at least 60% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 70% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 80% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 90% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 91% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 92% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 93% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 94% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 95% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 96% reduction of the targeted gene expression . In some embodiments, the base editing efficiency may result in at least 97% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 98% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 99% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in knockout (100% knockdown of the gene expression) of the gene that is targeted.

In some embodiments, any of the base editor systems provided herein result in less than 50%, less than 40%, less than 30%, less than 20%, less than 19%, less than 18%, less than 17%, less than 16%, less than 15%, less than 14%, less than 13%, less than 12%, less than 11%, less than 10%, less than 9%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, less than 1%, less than 0.9%, less than 0.8%, less than 0.7%, less than 0.6%, less than 0.5%, less than 0.4%, less than 0.3%, less than 0.2%, less than 0.1%, less than 0.09%, less than 0.08%, less than 0.07%, less than 0.06%, less than 0.05%, less than 0.04%, less than 0.03%, less than 0.02%, or less than 0.01% indel formation in the target polynucleotide sequence.

In some embodiments, targeted modifications, *e.g.*, single base editing, are used simultaneously to target at least 4, 5, 6, 7, 8, 9, 10, 11, 12 13, 14, 15, 16, 17 ,18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 different endogenous sequences for base editing with different guide RNAs. In some embodiments, targeted modifications, *e.g.* single base editing, are used to

sequentially target at least 4, 5, 6, 7, 8, 9, 10, 11, 12 13, 14, 15, 16, 17 ,18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 50, or more different endogenous gene sequences for base editing with different guide RNAs.

Some aspects of the disclosure are based on the recognition that any of the base editors provided herein are capable of efficiently generating an intended mutation, such as a point mutation, in a nucleic acid (*e.g.*, a nucleic acid within a genome of a subject) without generating a significant number of unintended mutations, such as unintended point mutations (*i.e.*, mutation of bystanders). In some embodiments, any of the base editors provided herein are capable of generating at least 0.01% of intended mutations (*i.e.*, at least 0.01% base editing efficiency). In some embodiments, any of the base editors provided herein are capable of generating at least 0.01%, 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, 45%, 50%, 60%, 70%, 80%, 90%, 95%, or 99% of intended mutations.

In some embodiments, any of the base editor systems comprising one of the ABE8 base editor variants described herein result in less than 50%, less than 40%, less than 30%, less than 20%, less than 19%, less than 18%, less than 17%, less than 16%, less than 15%, less than 14%, less than 13%, less than 12%, less than 11%, less than 10%, less than 9%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, less than 1%, less than 0.9%, less than 0.8%, less than 0.7%, less than 0.6%, less than 0.5%, less than 0.4%, less than 0.3%, less than 0.2%, less than 0.1%, less than 0.09%, less than 0.08%, less than 0.07%, less than 0.06%, less than 0.05%, less than 0.04%, less than 0.03%, less than 0.02%, or less than 0.01% indel formation in the target polynucleotide sequence. In some embodiments, any of the base editor systems comprising one of the ABE8 base editor variants described herein result in less than 0.8% indel formation in the target polynucleotide sequence. In some embodiments, any of the base editor systems comprising one of the ABE8 base editor variants described herein result in at most 0.8% indel formation in the target polynucleotide sequence. In some embodiments, any of the base editor systems comprising one of the ABE8 base editor variants described herein result in less than 0.3% indel formation in the target polynucleotide sequence. In some embodiments, any of the base editor systems comprising one of the ABE8 base editor variants described results in lower indel formation in the target polynucleotide sequence compared to a base editor system comprising one of ABE7 base editors. In some embodiments, any of the base editor systems comprising one of the ABE8 base editor variants described herein results in lower indel formation in the target polynucleotide sequence compared to a base editor system comprising an ABE7.10.

In some embodiments, any of the base editor systems comprising one of the ABE8 base editor variants described herein has reduction in indel frequency compared to a base editor system comprising one of the ABE7 base editors. In some embodiments, any of the base editor systems comprising one of the ABE8 base editor variants described herein has at least 0.01%, at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% reduction in indel frequency compared to a base editor system comprising one of the ABE7 base editors. In some embodiments, a base editor system comprising one of the ABE8 base editor variants described herein has at least 0.01%, at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% reduction in indel frequency compared to a base editor system comprising an ABE7.10.

The invention provides adenosine deaminase variants (*e.g.*, ABE8 variants) that have increased efficiency and specificity. In particular, the adenosine deaminase variants described herein are more likely to edit a desired base within a polynucleotide, and are less likely to edit bases that are not intended to be altered (*e.g.*, "bystanders").

In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced bystander editing or mutations. In some embodiments, an unintended editing or mutation is a bystander mutation or bystander editing, for example, base editing of a target base (*e.g.*, A or C) in an unintended or non-target position in a target window of a target nucleotide sequence. In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced bystander editing or mutations compared to a base editor system comprising an ABE7 base editor, *e.g.*, ABE7.10. In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced bystander editing or mutations by at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% compared to a base editor system comprising an ABE7 base editor, *e.g.*, ABE7.10. In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has

reduced bystander editing or mutations by at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, or at least 3.0 fold compared to a base editor system comprising an ABE7 base editor, *e.g.*, ABE7.10.

In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced spurious editing. In some embodiments, an unintended editing or mutation is a spurious mutation or spurious editing, for example, non-specific editing or guide independent editing of a target base (*e.g.*, A or C) in an unintended or non-target region of the genome. In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced spurious editing compared to a base editor system comprising an ABE7 base editor, *e.g.*, ABE7.10. In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced spurious editing by at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% compared to a base editor system comprising an ABE7 base editor, *e.g.*, ABE7.10. In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced spurious editing by at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, or at least 3.0 fold compared to a base editor system comprising an ABE7 base editor, *e.g.*, ABE7.10.

In some embodiments, any of the ABE8 base editor variants described herein have at least 0.01%, at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% base editing efficiency. In some embodiments, the base editing efficiency may be measured by calculating the percentage of edited nucleobases in a population of cells. In some embodiments, any of the ABE8 base editor variants described herein have base editing efficiency of at least 0.01%, at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least

25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% as measured by edited nucleobases in a population of cells.

In some embodiments, any of the ABE8 base editor variants described herein has higher base editing efficiency compared to the ABE7 base editors. In some embodiments, any of the ABE8 base editor variants described herein have at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, at least 100%, at least 105%, at least 110%, at least 115%, at least 120%, at least 125%, at least 130%, at least 135%, at least 140%, at least 145%, at least 150%, at least 155%, at least 160%, at least 165%, at least 170%, at least 175%, at least 180%, at least 185%, at least 190%, at least 195%, at least 200%, at least 210%, at least 220%, at least 230%, at least 240%, at least 250%, at least 260%, at least 270%, at least 280%, at least 290%, at least 300%, at least 310%, at least 320%, at least 330%, at least 340%, at least 350%, at least 360%, at least 370%, at least 380%, at least 390%, at least 400%, at least 450%, or at least 500% higher base editing efficiency compared to an ABE7 base editor, *e.g.*, ABE7.10.

In some embodiments, any of the ABE8 base editor variants described herein has at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, at least 3.0 fold, at least 3.1 fold, at least 3.2, at least 3.3 fold, at least 3.4 fold, at least 3.5 fold, at least 3.6 fold, at least 3.7 fold, at least 3.8 fold, at least 3.9 fold, at least 4.0 fold, at least 4.1 fold, at least 4.2 fold, at least 4.3 fold, at least 4.4 fold, at least 4.5 fold, at least 4.6 fold, at least 4.7 fold, at least 4.8 fold, at least 4.9 fold, or at least 5.0 fold higher base editing efficiency compared to an ABE7 base editor, *e.g.*, ABE7.10.

In some embodiments, any of the ABE8 base editor variants described herein have at least 0.01%, at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% on-target base editing efficiency. In some embodiments, any of the ABE8 base editor variants described herein have on-target base editing efficiency of at least 0.01%, at least 1%, at least 2%, at least 3%, at least 4%, at

least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% as measured by edited target nucleobases in a population of cells.

In some embodiments, any of the ABE8 base editor variants described herein has higher on-target base editing efficiency compared to the ABE7 base editors. In some embodiments, any of the ABE8 base editor variants described herein have at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, at least 100%, at least 105%, at least 110%, at least 115%, at least 120%, at least 125%, at least 130%, at least 135%, at least 140%, at least 145%, at least 150%, at least 155%, at least 160%, at least 165%, at least 170%, at least 175%, at least 180%, at least 185%, at least 190%, at least 195%, at least 200%, at least 210%, at least 220%, at least 230%, at least 240%, at least 250%, at least 260%, at least 270%, at least 280%, at least 290%, at least 300%, at least 310%, at least 320%, at least 330%, at least 340%, at least 350%, at least 360%, at least 370%, at least 380%, at least 390%, at least 400%, at least 450%, or at least 500% higher on-target base editing efficiency compared to an ABE7 base editor, *e.g.*, ABE7.10.

In some embodiments, any of the ABE8 base editor variants described herein has at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, at least 3.0 fold, at least 3.1 fold, at least 3.2 fold, at least 3.3 fold, at least 3.4 fold, at least 3.5 fold, at least 3.6 fold, at least 3.7 fold, at least 3.8 fold, at least 3.9 fold, at least 4.0 fold, at least 4.1 fold, at least 4.2 fold, at least 4.3 fold, at least 4.4 fold, at least 4.5 fold, at least 4.6 fold, at least 4.7 fold, at least 4.8 fold, at least 4.9 fold, or at least 5.0 fold higher on-target base editing efficiency compared to an ABE7 base editor, *e.g.*, ABE7.10.

The ABE8 base editor variants described herein may be delivered to a host cell via a plasmid, a vector, a LNP complex, or an mRNA. In some embodiments, any of the ABE8 base editor variants described herein is delivered to a host cell as an mRNA. In some embodiments, an ABE8 base editor delivered via a nucleic acid based delivery system, *e.g.*, an mRNA, has on-target editing efficiency of at least at least 1%, at least 2%, at least 3%, at

least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% as measured by edited nucleobases. In some embodiments, an ABE8 base editor delivered by an mRNA system has higher base editing efficiency compared to an ABE8 base editor delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described herein has at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, at least 100%, at least 105%, at least 110%, at least 115%, at least 120%, at least 125%, at least 130%, at least 135%, at least 140%, at least 145%, at least 150%, at least 155%, at least 160%, at least 165%, at least 170%, at least 175%, at least 180%, at least 185%, at least 190%, at least 195%, at least 200%, at least 210%, at least 220%, at least 230%, at least 240%, at least 250%, at least 260%, at least 270%, at least 280%, at least 290%, at least 300% higher, at least 310%, at least 320%, at least 330%, at least 340%, at least 350%, at least 360%, at least 370%, at least 380%, at least 390%, at least 400%, at least 450%, or at least 500% on-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described herein has at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, at least 3.0 fold, at least 3.1 fold, at least 3.2 fold, at least 3.3 fold, at least 3.4 fold, at least 3.5 fold, at least 3.6 fold, at least 3.7 fold, at least 3.8 fold, at least 3.9 fold, at least 4.0 fold, at least 4.1 fold, at least 4.2 fold, at least 4.3 fold, at least 4.4 fold, at least 4.5 fold, at least 4.6 fold, at least 4.7 fold, at least 4.8 fold, at least 4.9 fold, or at least 5.0 fold higher on-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system.

In some embodiments, any of the base editor systems comprising one of the ABE8 base editor variants described herein result in less than 50%, less than 40%, less than 30%, less than 20%, less than 19%, less than 18%, less than 17%, less than 16%, less than 15%, less than 14%, less than 13%, less than 12%, less than 11%, less than 10%, less than 9%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, less than 1%, less than 0.9%, less than 0.8%, less than 0.7%, less than 0.6%, less than 0.5%,

less than 0.4%, less than 0.3%, less than 0.2%, less than 0.1%, less than 0.09%, less than 0.08%, less than 0.07%, less than 0.06%, less than 0.05%, less than 0.04%, less than 0.03%, less than 0.02%, or less than 0.01% off-target editing in the target polynucleotide sequence.

In some embodiments, any of the ABE8 base editor variants described herein has lower guided off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described herein has at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% lower guided off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described herein has at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, or at least 3.0 fold lower guided off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described herein has at least about 2.2 fold decrease in guided off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system.

In some embodiments, any of the ABE8 base editor variants described herein has lower guide-independent off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described herein has at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% lower guide-independent off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described herein has at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold,

at least 3.0 fold, at least 5.0 fold, at least 10.0 fold, at least 20.0 fold, at least 50.0 fold, at least 70.0 fold, at least 100.0 fold, at least 120.0 fold, at least 130.0 fold, or at least 150.0 fold lower guide-independent off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, ABE8 base editor variants described herein has 134.0 fold decrease in guide-independent off-target editing efficiency (*e.g.*, spurious RNA deamination) when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, ABE8 base editor variants described herein does not increase guide-independent mutation rates across the genome.

In some embodiments, a single gene delivery event (*e.g.*, by transduction, transfection, electroporation or any other method) can be used to target base editing of 5 sequences within a cell's genome. In some embodiments, a single gene delivery event can be used to target base editing of 6 sequences within a cell's genome. In some embodiments, a single gene delivery event can be used to target base editing of 7 sequences within a cell's genome. In some embodiments, a single electroporation event can be used to target base editing of 8 sequences within a cell's genome. In some embodiments, a single gene delivery event can be used to target base editing of 9 sequences within a cell's genome. In some embodiments, a single gene delivery event can be used to target base editing of 10 sequences within a cell's genome. In some embodiments, a single gene delivery event can be used to target base editing of 20 sequences within a cell's genome. In some embodiments, a single gene delivery event can be used to target base editing of 30 sequences within a cell's genome. In some embodiments, a single gene delivery event can be used to target base editing of 40 sequences within a cell's genome. In some embodiments, a single gene delivery event can be used to target base editing of 50 sequences within a cell's genome.

In some embodiments, the method described herein, for example, the base editing methods has minimum to no off-target effects.

In some embodiments, the base editing method described herein results in at least 50% of a cell population that have been successfully edited (*i.e.*, cells that have been successfully engineered). In some embodiments, the base editing method described herein results in at least 55% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least 60% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least 65% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least

70% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least 75% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least 80% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least 85% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least 90% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least 95% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in about 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% of a cell population that have been successfully edited.

In some embodiments, the live cell recovery following a base editing intervention is greater than at least 60%, 70%, 80%, 90% of the starting cell population at the time of the base editing event. In some embodiments, the live cell recovery as described above is about 70%. In some embodiments, the live cell recovery as described above is about 75%. In some embodiments, the live cell recovery as described above is about 80%. In some embodiments, the live cell recovery as described above is about 85%. In some embodiments, the live cell recovery as described above is about 90%, or about 91%, 92%, 93%, 94% 95%, 96%, 97%, 98%, or 99%, or 100% of the cells in the population at the time of the base editing event.

In some embodiments the engineered cell population can be further expanded *in vitro* by about 2 fold, about 3-fold, about 4-fold, about 5-fold, about 6-fold, about 7-fold, about 8-fold, about 9-fold, about 10-fold, about 15-fold, about 20-fold, about 25-fold, about 30-fold, about 35-fold, about 40-fold, about 45-fold, about 50-fold, or about 100-fold.

The number of intended mutations and indels can be determined using any suitable method, for example, as described in International PCT Application Nos. PCT/2017/045381 (WO2018/027078) and PCT/US2016/058344 (WO2017/070632); Komor, A.C., et al., "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., et al., "Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); and Komor, A.C., et al., "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017); the entire contents of which are hereby incorporated by reference.

In some embodiments, to calculate indel frequencies, sequencing reads are scanned for exact matches to two 10-bp sequences that flank both sides of a window in which indels can occur. If no exact matches are located, the read is excluded from analysis. If the length of this indel window exactly matches the reference sequence the read is classified as not containing an indel. If the indel window is two or more bases longer or shorter than the reference sequence, then the sequencing read is classified as an insertion or deletion, respectively. In some embodiments, the base editors provided herein can limit formation of indels in a region of a nucleic acid. In some embodiments, the region is at a nucleotide targeted by a base editor or a region within 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides of a nucleotide targeted by a base editor.

The number of indels formed at a target nucleotide region can depend on the amount of time a nucleic acid (*e.g.*, a nucleic acid within the genome of a cell) is exposed to a base editor. In some embodiments, the number or proportion of indels is determined after at least 1 hour, at least 2 hours, at least 6 hours, at least 12 hours, at least 24 hours, at least 36 hours, at least 48 hours, at least 3 days, at least 4 days, at least 5 days, at least 7 days, at least 10 days, or at least 14 days of exposing the target nucleotide sequence (*e.g.*, a nucleic acid within the genome of a cell) to a base editor. It should be appreciated that the characteristics of the base editors as described herein can be applied to any of the fusion proteins, or methods of using the fusion proteins provided herein.

Details of base editor efficiency are described in International PCT Application Nos. PCT/2017/045381 (WO 2018/027078) and PCT/US2016/058344 (WO 2017/070632), each of which is incorporated herein by reference for its entirety. Also see Komor, A.C., et al., "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., et al., "Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); and Komor, A.C., et al., "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017), the entire contents of which are hereby incorporated by reference. In some embodiments, editing of a plurality of nucleobase pairs in one or more genes using the methods provided herein results in formation of at least one intended mutation. In some embodiments, said formation of said at least one intended mutation results in the disruption the normal function of a gene. In some embodiments, said formation of said at least one intended mutation results decreases or eliminates the expression of a protein

encoded by a gene. It should be appreciated that multiplex editing can be accomplished using any method or combination of methods provided herein.

Multiplex Editing

In some embodiments, the base editor system provided herein is capable of multiplex editing of a plurality of nucleobase pairs in one or more genes or polynucleotide sequences. In some embodiments, the plurality of nucleobase pairs is located in the same gene or in one or more genes, wherein at least one gene is located in a different locus. In some embodiments, the multiplex editing can comprise one or more guide polynucleotides. In some embodiments, the multiplex editing can comprise one or more base editor systems. In some embodiments, the multiplex editing can comprise one or more base editor systems with a single guide polynucleotide or a plurality of guide polynucleotides. In some embodiments, the multiplex editing can comprise one or more guide polynucleotides with a single base editor system. In some embodiments, the multiplex editing can comprise at least one guide polynucleotide that does or does not require a PAM sequence to target binding to a target polynucleotide sequence. In some embodiments, the multiplex editing can comprise a mix of at least one guide polynucleotide that does not require a PAM sequence to target binding to a target polynucleotide sequence and at least one guide polynucleotide that require a PAM sequence to target binding to a target polynucleotide sequence. It should be appreciated that the characteristics of the multiplex editing using any of the base editors as described herein can be applied to any combination of methods using any base editor provided herein. It should also be appreciated that the multiplex editing using any of the base editors as described herein can comprise a sequential editing of a plurality of nucleobase pairs.

In some embodiments, the plurality of nucleobase pairs are in one more genes. In some embodiments, the plurality of nucleobase pairs is in the same gene. In some embodiments, at least one gene in the one more genes is located in a different locus.

In some embodiments, the plurality of nucleobase pairs are in one or more target polynucleotide sequences. In some embodiments, the plurality of nucleobase pairs is in the same target polynucleotide sequence. In some embodiments, the one or more target polynucleotide sequences is present in the intron of a polynucleotide encoding a self-inactivating base editor.

In some embodiments, the editing is editing of the plurality of nucleobase pairs in at least one protein coding region, in at least one protein non-coding region, or in at least one protein coding region and at least one protein non-coding region.

In some embodiments, the editing is in conjunction with one or more guide polynucleotides. In some embodiments, the base editor system can comprise one or more base editor systems. In some embodiments, the base editor system can comprise one or more base editor systems in conjunction with a single guide polynucleotide or a plurality of guide polynucleotides. In some embodiments, the editing is in conjunction with one or more guide polynucleotide with a single base editor system. In some embodiments, the editing is in conjunction with at least one guide polynucleotide that does not require a PAM sequence to target binding to a target polynucleotide sequence or with at least one guide polynucleotide that requires a PAM sequence to target binding to a target polynucleotide sequence, or with a mix of at least one guide polynucleotide that does not require a PAM sequence to target binding to a target polynucleotide sequence and at least one guide polynucleotide that does require a PAM sequence to target binding to a target polynucleotide sequence. It should be appreciated that the characteristics of the multiplex editing using any of the base editors as described herein can be applied to any of combination of the methods of using any of the base editors provided herein. It should also be appreciated that the editing can comprise a sequential editing of a plurality of nucleobase pairs.

In some embodiments, the base editor system capable of multiplex editing of a plurality of nucleobase pairs in one or more genes comprises one of ABE7, ABE8, and/or ABE9 base editors. In some embodiments, the base editor system capable of multiplex editing comprising one of the ABE8 base editor variants described herein has higher multiplex editing efficiency compared to the base editor system capable of multiplex editing comprising one of ABE7 base editors. In some embodiments, the base editor system capable of multiplex editing comprising one of the ABE8 base editor variants described herein has at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, at least 100%, at least 105%, at least 110%, at least 115%, at least 120%, at least 125%, at least 130%, at least 135%, at least 140%, at least 145%, at least 150%, at least 155%, at least 160%, at least 165%, at least 170%, at least 175%, at least 180%, at least 185%, at least 190%, at least 195%, at least 200%, at least 210%, at least 220%, at least 230%, at least 240%, at least 250%, at least 260%, at least 270%, at least

280%, at least 290%, at least 300% higher, at least 310%, at least 320%, at least 330%, at least 340%, at least 350%, at least 360%, at least 370%, at least 380%, at least 390%, at least 400%, at least 450%, or at least 500% higher multiplex editing efficiency compared the base editor system capable of multiplex editing comprising one of ABE7 base editors. In some embodiments, the base editor system capable of multiplex editing comprising one of the ABE8 base editor variants described herein has at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, at least 3.0 fold, at least 3.1 fold, at least 3.2 fold, at least 3.3 fold, at least 3.4 fold, at least 3.5 fold, at least 4.0 fold, at least 4.5 fold, at least 5.0 fold, at least 5.5 fold, or at least 6.0 fold higher multiplex editing efficiency compared the base editor system capable of multiplex editing comprising one of ABE7 base editors.

DELIVERY SYSTEM

The suitability of nucleobase editors to target one or more nucleotides in a polynucleotide sequence (*e.g.*, a gene or intron) is evaluated as described herein. In one embodiment, a single cell of interest is transfected, transduced, or otherwise modified with a nucleic acid molecule or molecules encoding a base editing system described herein together with a small amount of a vector encoding a reporter (*e.g.*, GFP). These cells can be any cell line known in the art (*e.g.*, HEK293T cells). Alternatively, primary cells (*e.g.*, human) may be used. Cells may also be obtained from a subject or individual, such as from tissue biopsy, surgery, blood, plasma, serum, or other biological fluid. Such cells may be relevant to the eventual cell target.

Delivery may be performed using a viral vector. In one embodiment, transfection may be performed using lipid transfection (such as Lipofectamine or Fugene) or by electroporation. Following transfection, expression of a reporter (*e.g.*, GFP) can be determined either by fluorescence microscopy or by flow cytometry to confirm consistent and high levels of transfection. These preliminary transfections can comprise different nucleobase editors to determine which combinations of editors give the greatest activity. The system can comprise one or more different vectors. In one embodiment, the base editor is codon optimized for expression of the desired cell type, preferentially a eukaryotic cell, preferably a mammalian cell or a human cell.

The activity of the nucleobase editor is assessed as described herein, *i.e.*, by sequencing the genome of the cells to detect alterations in a target sequence. For Sanger sequencing, purified PCR amplicons are cloned into a plasmid backbone, transformed, miniprepped and sequenced with a single primer. Sequencing may also be performed using next generation sequencing (NGS) techniques. When using next generation sequencing, amplicons may be 300-500 bp with the intended cut site placed asymmetrically. Following PCR, next generation sequencing adapters and barcodes (for example Illumina multiplex adapters and indexes) may be added to the ends of the amplicon, *e.g.*, for use in high throughput sequencing (for example on an Illumina MiSeq). The fusion proteins that induce the greatest levels of target specific alterations in initial tests can be selected for further evaluation.

In particular embodiments, the nucleobase editors are used to target polynucleotides of interest. In one embodiment, a nucleobase editor of the invention is delivered to cells in conjunction with one or more guide RNAs that are used to target one or more nucleic acid sequences of interest within the genome of a cell, thereby altering the target gene(s). In some embodiments, a base editor is targeted by one or more guide RNAs to introduce one or more edits to the sequence of one or more genes of interest. In some embodiments, the one or more edits to the sequence of one or more genes of interest decrease or eliminate expression of the protein encoded by the gene in the host cell. In some embodiments, expression of one or more proteins encoded by one or more genes of interest is completely knocked out or eliminated in the host cell.

In some embodiments, a nucleobase editor or a polynucleotide encoding a nucleobase editor of the invention is delivered to cells (*e.g.*, host cells) in conjunction with one or more guide RNAs that target a heterologous intron within the polynucleotide sequence encoding the base editor, thereby altering the targeted intron (*e.g.*, splice acceptor, splice donor site). In some embodiments, the one or more edits to the sequence of the intron decreases or eliminates the expression, activity, or level of base editing activity

In some embodiments, the host cell is selected from a bacterial cell, plant cell, insect cell, human cell, or mammalian cell. In some embodiments, the host cell is a mammalian cell. In some embodiments, the host cell is a human cell. In some embodiments, the cell is *in vitro*. In some embodiments, the cell is *in vivo*.

Nucleic Acid-Based Delivery of Base Editor Systems

Nucleic acid molecules encoding a base editor system according to the present disclosure can be administered to subjects or delivered into cells *in vitro* or *in vivo* by art-known methods or as described herein. In some embodiments, a nucleic acid molecule encoding a self-inactivating base editor includes an intron that can be edited to reduce the level, expression, or activity of the base editor in a cell. For example, a base editor system comprising a deaminase (*e.g.*, cytidine or adenine deaminase) can be delivered by vectors (*e.g.*, viral or non-viral vectors), or by naked DNA, DNA complexes, lipid nanoparticles, or a combination of the aforementioned compositions.

Nanoparticles, which can be organic or inorganic, are useful for delivering a base editor system or component thereof. Nanoparticles are well known in the art and any suitable nanoparticle can be used to deliver a base editor system or component thereof, or a nucleic acid molecule encoding such components. In one example, organic (*e.g.* lipid and/or polymer) nanoparticles are suitable for use as delivery vehicles in certain embodiments of this disclosure. Exemplary lipids for use in nanoparticle formulations, and/or gene transfer are shown in Table 16 (below).

Table 16

Lipids Used for Gene Transfer		
Lipid	Abbreviation	Feature
1,2-Dioleoyl-sn-glycero-3-phosphatidylcholine	DOPC	Helper
1,2-Dioleoyl-sn-glycero-3-phosphatidylethanolamine	DOPE	Helper
Cholesterol		Helper
N-[1-(2,3-Dioleyloxy)prophyl]N,N,N-trimethylammonium chloride	DOTMA	Cationic
1,2-Dioleoyloxy-3-trimethylammonium-propane	DOTAP	Cationic
Dioctadecylamidoglycylspermine	DOGS	Cationic
N-(3-Aminopropyl)-N,N-dimethyl-2,3-bis(dodecyloxy)-1-propanaminium bromide	GAP-DLRIE	Cationic
Cetyltrimethylammonium bromide	CTAB	Cationic
6-Lauroxyhexyl ornithinate	LHON	Cationic
1-(2,3-Dioleoyloxypropyl)-2,4,6-trimethylpyridinium	2Oc	Cationic
2,3-Dioleyloxy-N-[2(sperminecarboxamido-ethyl]-N,N-dimethyl-1-propanaminium trifluoroacetate	DOSPA	Cationic
1,2-Dioleyl-3-trimethylammonium-propane	DOPA	Cationic
N-(2-Hydroxyethyl)-N,N-dimethyl-2,3-bis(tetradecyloxy)-1-propanaminium bromide	MDRIE	Cationic
Dimyristooxypropyl dimethyl hydroxyethyl ammonium bromide	DMRI	Cationic
3β-[N-(N',N'-Dimethylaminoethane)-carbamoyl]cholesterol	DC-Chol	Cationic
Bis-guanidium-tren-cholesterol	BGTC	Cationic
1,3-Diodeoxy-2-(6-carboxy-spermyl)-propylamide	DOSPER	Cationic

Lipids Used for Gene Transfer		
Lipid	Abbreviation	Feature
Dimethyloctadecylammonium bromide	DDAB	Cationic
Dioctadecylamidoglicylspermidin	DSL	Cationic
rac-[(2,3-Dioctadecyloxypropyl)(2-hydroxyethyl)]-dimethylammonium chloride	CLIP-1	Cationic
rac-[2(2,3-Dihexadecyloxypropyl-oxymethyloxy)ethyl]trimethylammoniun bromide	CLIP-6	Cationic
Ethyldimyristoylphosphatidylcholine	EDMPC	Cationic
1,2-Distearyloxy-N,N-dimethyl-3-aminopropane	DSDMA	Cationic
1,2-Dimyristoyl-trimethylammonium propane	DMTAP	Cationic
O,O'-Dimyristyl-N-lysyl aspartate	DMKE	Cationic
1,2-Distearoyl-sn-glycero-3-ethylpho sphocholine	DSEPC	Cationic
N-Palmitoyl D-erythro-sphingosyl carbamoyl-spermine	CCS	Cationic
N-t-Butyl-N0-tetradecyl-3-tetradecylaminopropionamidine	diC14-amidine	Cationic
Octadecenolyoxy[ethyl-2-heptadecenyl-3 hydroxyethyl] imidazolinium chloride	DOTIM	Cationic
N1 -Cholesteryloxycarbonyl-3,7-diazanonane-1,9-diamine	CDAN	Cationic
2-(3-[Bis(3-amino-propyl)-amino]propylamino)-N-ditetradecylcarbamoylme-ethyl-acetamide	RPR209120	Cationic
1,2-dilinoleyloxy-3-dimethylaminopropane	DLinDMA	Cationic
2,2-dilinoleyl-4-dimethylaminoethyl-[1,3]-dioxolane	DLin-KC2-DMA	Cationic
dilinoleyl-methyl-4-dimethylaminobutyrate	DLin-MC3-DMA	Cationic

Table 17 lists exemplary polymers for use in gene transfer and/or nanoparticle formulations.

Table 17

Polymers Used for Gene Transfer	
Polymer	Abbreviation
Poly(ethylene)glycol	PEG
Polyethylenimine	PEI
Dithiobis (succinimidylpropionate)	DSP
Dimethyl-3,3'-dithiobispropionimidate	DTBP
Poly(ethylene imine)biscarbamate	PEIC
Poly(L-lysine)	PLL
Histidine modified PLL	
Poly(N-vinylpyrrolidone)	PVP
Poly(propylenimine)	PPI
Poly(amidoamine)	PAMAM
Poly(amidoethylenimine)	SS-PAEI
Triethylenetetramine	TETA
Poly(β-aminoester)	
Poly(4-hydroxy-L-proline ester)	PHP
Poly(allylamine)	
Poly(α-[4-aminobutyl]-L-glycolic acid)	PAGA
Poly(D,L-lactic-co-glycolic acid)	PLGA
Poly(N-ethyl-4-vinylpyridinium bromide)	

Polymers Used for Gene Transfer	
Polymer	Abbreviation
Poly(phosphazene)s	PPZ
Poly(phosphoester)s	PPE
Poly(phosphoramidate)s	PPA
Poly(N-2-hydroxypropylmethacrylamide)	pHPMA
Poly (2-(dimethylamino)ethyl methacrylate)	pDMAEMA
Poly(2-aminoethyl propylene phosphate)	PPE-EA
Chitosan	
Galactosylated chitosan	
N-Dodacylated chitosan	
Histone	
Collagen	
Dextran-spermine	D-SPM

Table 18 summarizes delivery methods for a polynucleotide encoding a fusion protein described herein.

Table 18

Delivery	Vector/Mode	Delivery into Non-Dividing Cells	Duration of Expression	Genome Integration	Type of Molecule Delivered
Physical	(*e.g.*, electroporation, particle gun, Calcium Phosphate transfection	YES	Transient	NO	Nucleic Acids and Proteins
Viral	Retrovirus	NO	Stable	YES	RNA
	Lentivirus	YES	Stable	YES/NO with modification	RNA
	Adenovirus	YES	Transient	NO	DNA
	Adeno-Associated Virus (AAV)	YES	Stable	NO	DNA
	Vaccinia Virus	YES	Very Transient	NO	DNA
	Herpes Simplex Virus	YES	Stable	NO	DNA
Non-Viral	Cationic Liposomes	YES	Transient	Depends on what is delivered	Nucleic Acids and Proteins
	Polymeric Nanoparticles	YES	Transient	Depends on what is delivered	Nucleic Acids and Proteins
Biological Non-Viral	Attenuated Bacteria	YES	Transient	NO	Nucleic Acids

Delivery	Vector/Mode	Delivery into Non-Dividing Cells	Duration of Expression	Genome Integration	Type of Molecule Delivered
Delivery Vehicles	Engineered Bacteriophages	YES	Transient	NO	Nucleic Acids
	Mammalian Virus-like Particles	YES	Transient	NO	Nucleic Acids
	Biological liposomes: Erythrocyte Ghosts and Exosomes	YES	Transient	NO	Nucleic Acids

In another aspect, the delivery of base editor system components or nucleic acids encoding such components, for example, a polynucleotide programmable nucleotide binding domain (*e.g.*, Cas9) such as, for example, Cas9 or variants thereof, and a gRNA targeting a nucleic acid sequence of interest, may be accomplished by delivering the ribonucleoprotein (RNP) to cells. The RNP comprises a polynucleotide programmable nucleotide binding domain (*e.g.*, Cas9), in complex with the targeting gRNA. RNPs or polynucleotides described herein may be delivered to cells using known methods, such as electroporation, nucleofection, or cationic lipid-mediated methods, for example, as reported by Zuris, J.A. et al., 2015, Nat. Biotechnology, 33(1):73-80, which is incorporated by reference in its entirety. RNPs are advantageous for use in CRISPR base editing systems, particularly for cells that are difficult to transfect, such as primary cells. In addition, RNPs can also alleviate difficulties that may occur with protein expression in cells, especially when eukaryotic promoters, *e.g.*, CMV or EF1A, which may be used in CRISPR plasmids, are not well-expressed. Advantageously, the use of RNPs does not require the delivery of foreign DNA into cells. Moreover, because an RNP comprising a nucleic acid binding protein and gRNA complex is degraded over time, the use of RNPs has the potential to limit off-target effects. In a manner similar to that for plasmid based techniques, RNPs can be used to deliver binding protein (*e.g.*, Cas9 variants) and to direct homology directed repair (HDR).

Nucleic acid molecules encoding a base editor system can be delivered directly to cells as naked DNA or RNA by means of transfection or electroporation, for example, or can be conjugated to molecules (*e.g.*, N-acetylgalactosamine) promoting uptake by the target cells. Vectors encoding base editor systems and/or their components can also be used. In particular embodiments, a polynucleotide, *e.g.* a mRNA encoding a base editor system or a

functional component thereof, may be co-electroporated with one or more guide RNAs as described herein.

Nucleic acid vectors can comprise one or more sequences encoding a domain of a fusion protein described herein. A vector can also encode a protein component of a base editor system operably linked to a nuclear localization signal, nucleolar localization signal, or mitochondrial localization signal. As one example, a vector can include a Cas9 coding sequence that includes one or more nuclear localization sequences (*e.g.*, a nuclear localization sequence from SV40), and one or more deaminases.

The vector can also include any suitable number of regulatory/control elements, *e.g.*, promoters, enhancers, introns, polyadenylation signals, Kozak consensus sequences, or internal ribosome entry sites (IRES). These elements are well known in the art.

Vectors according to this disclosure include recombinant viral vectors. Exemplary viral vectors are set forth herein above. Other viral vectors known in the art can also be used. In addition, viral particles can be used to deliver base editor system components in nucleic acid and/or protein form. For example, "empty" viral particles can be assembled to contain a base editor system or component as cargo. Viral vectors and viral particles can also be engineered to incorporate targeting ligands to alter target tissue specificity.

Vectors described herein may comprise regulatory elements to drive expression of a base editor system or component thereof. Such vectors include adeno-associated viruses with inverted long terminal repeats (AAV ITR). The use of AAV-ITR can be advantageous for eliminating the need for an additional promoter element, which can take up space in the vector. The additional space freed up can be used to drive the expression of additional elements, such as a guide nucleic acid or a selectable marker. ITR activity can be used to reduce potential toxicity due to over expression.

Any suitable promoter can be used to drive expression of a base editor system or component thereof and, where appropriate, the guide nucleic acid. For ubiquitous expression, promoters include CMV, CAG, CBh, PGK, SV40, Ferritin heavy or light chains. For brain or other CNS cell expression, suitable promoters include: SynapsinI for all neurons, CaMKIIalpha for excitatory neurons, GAD67 or GAD65 or VGAT for GABAergic neurons. For liver cell expression, suitable promoters include the Albumin promoter. For lung cell expression, suitable promoters include SP-B. For endothelial cells, suitable promoters include ICAM. For hematopoietic cell expression suitable promoters include IFNbeta or CD45. For osteoblast expression suitable promoters can include OG-2.

In some embodiments, a base editor system of the present disclosure is of small enough size to allow separate promoters to drive expression of the base editor and a compatible guide nucleic acid within the same nucleic acid molecule. For instance, a vector or viral vector can comprise a first promoter operably linked to a nucleic acid encoding the base editor and a second promoter operably linked to the guide nucleic acid.

The promoter used to drive expression of a guide nucleic acid can include: Pol III promoters, such as U6 or H1 Use of Pol II promoter and intronic cassettes to express gRNA Adeno Associated Virus (AAV).

In particular embodiments, a fusion protein of the invention is encoded by a polynucleotide present in a viral vector (*e.g.*, adeno-associated virus (AAV), AAV3, AAV3b, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAVrh8, AAV10, and variants thereof), or a suitable capsid protein of any viral vector. Thus, in some aspects, the disclosure relates to the viral delivery of a fusion protein. Examples of viral vectors include retroviral vectors (*e.g.* Maloney murine leukemia virus, MML-V), adenoviral vectors (*e.g.* AD100), lentiviral vectors (HIV and FIV-based vectors), herpesvirus vectors (*e.g.* HSV-2).

Viral Vectors

A base editor described herein can be delivered with a viral vector. In some embodiments, a base editor disclosed herein can be encoded on a nucleic acid that is contained in a viral vector. In some embodiments, one or more components of the base editor system can be encoded on one or more viral vectors. For example, a base editor and guide nucleic acid can be encoded on a single viral vector. In other embodiments, the base editor and guide nucleic acid are encoded on different viral vectors. In either case, the base editor and guide nucleic acid can each be operably linked to a promoter and terminator. The combination of components encoded on a viral vector can be determined by the cargo size constraints of the chosen viral vector.

The use of RNA or DNA viral based systems for the delivery of a base editor takes advantage of highly evolved processes for targeting a virus to specific cells in culture or in the host and trafficking the viral payload to the nucleus or host cell genome. Viral vectors can be administered directly to cells in culture, patients (*in vivo*), or they can be used to treat cells *in vitro*, and the modified cells can optionally be administered to patients (*ex vivo*). Conventional viral based systems could include retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods,

often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

Viral vectors can include lentivirus (*e.g.*, HIV and FIV-based vectors), Adenovirus (*e.g.*, AD100), Retrovirus (*e.g.*, Maloney murine leukemia virus, MML-V), herpesvirus vectors (*e.g.*, HSV-2), and Adeno-associated viruses (AAVs), or other plasmid or viral vector types, in particular, using formulations and doses from, for example, U.S. Patent No. 8,454,972 (formulations, doses for adenovirus), U.S. Patent No. 8,404,658 (formulations, doses for AAV) and U.S. Patent No. 5,846,946 (formulations, doses for DNA plasmids) and from clinical trials and publications regarding the clinical trials involving lentivirus, AAV and adenovirus. For example, for AAV, the route of administration, formulation and dose can be as in U.S. Patent No. 8,454,972 and as in clinical trials involving AAV. For Adenovirus, the route of administration, formulation and dose can be as in U.S. Patent No. 8,404,658 and as in clinical trials involving adenovirus. For plasmid delivery, the route of administration, formulation and dose can be as in U.S. Patent No. 5,846,946 and as in clinical studies involving plasmids. Doses can be based on or extrapolated to an average 70 kg individual (*e.g.* a male adult human), and can be adjusted for patients, subjects, mammals of different weight and species. Frequency of administration is within the ambit of the medical or veterinary practitioner (*e.g.*, physician, veterinarian), depending on usual factors including the age, sex, general health, other conditions of the patient or subject and the particular condition or symptoms being addressed. The viral vectors can be injected into the tissue of interest. For cell-type specific base editing, the expression of the base editor and optional guide nucleic acid can be driven by a cell-type specific promoter.

The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system would therefore depend on the target tissue. Retroviral vectors are comprised of cis-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum cis-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immuno deficiency virus (SIV), human immuno deficiency virus (HIV), and combinations thereof (See, *e.g.*, Buchscher et al., J. Virol. 66:2731-2739 (1992); Johann et al., J. Virol. 66:1635-1640 (1992); Sommnerfelt et al., Virol. 176:58-59 (1990); Wilson et al.,

J. Virol. 63:2374-2378 (1989); Miller et al., J. Virol. 65:2220-2224 (1991); PCT/US94/05700).

Retroviral vectors, especially lentiviral vectors, can require polynucleotide sequences smaller than a given length for efficient integration into a target cell. For example, retroviral vectors of length greater than 9 kb can result in low viral titers compared with those of smaller size. In some aspects, a base editor of the present disclosure is of sufficient size so as to enable efficient packaging and delivery into a target cell via a retroviral vector. In some embodiments, a base editor is of a size so as to allow efficient packing and delivery even when expressed together with a guide nucleic acid and/or other components of a targetable nuclease system.

Packaging cells are typically used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and psi.2 cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by producing a cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the polynucleotide(s) to be expressed. The missing viral functions are typically supplied in trans by the packaging cell line. For example, Adeno-associated virus ("AAV") vectors used in gene therapy typically only possess ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA can be packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely rep and cap, but lacking ITR sequences. The cell line can also be infected with adenovirus as a helper. The helper virus can promote replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid in some cases is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, *e.g.*, heat treatment to which adenovirus is more sensitive than AAV.

In applications where transient expression is preferred, adenoviral based systems can be used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. AAV vectors can also be used to transduce cells with target nucleic acids, *e.g.*, in the *in vitro* production of nucleic acids and peptides, and for *in vivo* and *ex vivo* gene therapy procedures (See, *e.g.*, West et al., Virology 160:38-47 (1987); U.S. Patent No. 4,797,368; WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J.

Clin. Invest. 94:1351 (1994). The construction of recombinant AAV vectors is described in a number of publications, including U.S. Patent No. 5,173,414; Tratschin et al., Mol. Cell. Biol. 5:3251-3260 (1985); Tratschin, et al., Mol. Cell. Biol. 4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); and Samulski et al., J. Virol. 63:03822-3828 (1989).

In some embodiments, AAV vectors are used to transduce a cell of interest with a polynucleotide encoding a base editor or base editor system as provided herein. AAV is a small, single-stranded DNA dependent virus belonging to the parvovirus family. The 4.7 kb wild-type (wt) AAV genome is made up of two genes that encode four replication proteins and three capsid proteins, respectively, and is flanked on either side by 145-bp inverted terminal repeats (ITRs). The virion is composed of three capsid proteins, Vp1, Vp2, and Vp3, produced in a 1:1:10 ratio from the same open reading frame but from differential splicing (Vp1) and alternative translational start sites (Vp2 and Vp3, respectively). Vp3 is the most abundant subunit in the virion and participates in receptor recognition at the cell surface defining the tropism of the virus. A phospholipase domain, which functions in viral infectivity, has been identified in the unique N terminus of Vp1.

Similar to wt AAV, recombinant AAV (rAAV) utilizes the cis-acting 145-bp ITRs to flank vector transgene cassettes, providing up to 4.5 kb for packaging of foreign DNA. Subsequent to infection, rAAV can express a fusion protein of the invention and persist without integration into the host genome by existing episomally in circular head-to-tail concatemers. Although there are numerous examples of rAAV success using this system, *in vitro* and *in vivo*, the limited packaging capacity has limited the use of AAV-mediated gene delivery when the length of the coding sequence of the gene is equal or greater in size than the wt AAV genome.

Viral vectors can be selected based on the application. For example, for *in vivo* gene delivery, AAV can be advantageous over other viral vectors. In some embodiments, AAV allows low toxicity, which can be due to the purification method not requiring ultra-centrifugation of cell particles that can activate the immune response. In some embodiments, AAV allows low probability of causing insertional mutagenesis because it doesn't integrate into the host genome. Adenoviruses are commonly used as vaccines because of the strong immunogenic response they induce. Packaging capacity of the viral vectors can limit the size of the base editor that can be packaged into the vector.

AAV has a packaging capacity of about 4.5 Kb or 4.75 Kb including two 145 base inverted terminal repeats (ITRs). This means disclosed base editor as well as a promoter and transcription terminator can fit into a single viral vector. Constructs larger than 4.5 or 4.75

Kb can lead to significantly reduced virus production. For example, SpCas9 is quite large, the gene itself is over 4.1 Kb, which makes it difficult for packing into AAV. Therefore, embodiments of the present disclosure include utilizing a disclosed base editor which is shorter in length than conventional base editors. In some examples, the base editors are less than 4 kb. Disclosed base editors can be less than 4.5 kb, 4.4 kb, 4.3 kb, 4.2 kb, 4.1 kb, 4 kb, 3.9 kb, 3.8 kb, 3.7 kb, 3.6 kb, 3.5 kb, 3.4 kb, 3.3 kb, 3.2 kb, 3.1 kb, 3 kb, 2.9 kb, 2.8 kb, 2.7 kb, 2.6 kb, 2.5 kb, 2 kb, or 1.5 kb. In some embodiments, the disclosed base editors are 4.5 kb or less in length.

An AAV can be AAV1, AAV2, AAV5 or any combination thereof. One can select the type of AAV with regard to the cells to be targeted; *e.g.*, one can select AAV serotypes 1, 2, 5 or a hybrid capsid AAV1, AAV2, AAV5 or any combination thereof for targeting brain or neuronal cells; and one can select AAV4 for targeting cardiac tissue. AAV8 is useful for delivery to the liver. A tabulation of certain AAV serotypes as to these cells can be found in Grimm, D. et al, J. Virol. 82: 5887-5911 (2008)).

In some embodiments, lentiviral vectors are used to transduce a cell of interest with a polynucleotide encoding a base editor or base editor system as provided herein. Lentiviruses are complex retroviruses that have the ability to infect and express their genes in both mitotic and post-mitotic cells. The most commonly known lentivirus is the human immunodeficiency virus (HIV), which uses the envelope glycoproteins of other viruses to target a broad range of cell types.

Lentiviruses can be prepared as follows. After cloning pCasES10 (which contains a lentiviral transfer plasmid backbone), HEK293FT at low passage (p=5) were seeded in a T-75 flask to 50% confluence the day before transfection in DMEM with 10% fetal bovine serum and without antibiotics. After 20 hours, media is changed to OptiMEM (serum-free) media and transfection was done 4 hours later. Cells are transfected with 10 μg of lentiviral transfer plasmid (pCasES10) and the following packaging plasmids: 5 μg of pMD2.G (VSV-g pseudotype), and 7.5 μg of psPAX2 (gag/pol/rev/tat). Transfection can be done in 4 mL OptiMEM with a cationic lipid delivery agent (50 μl Lipofectamine 2000 and 100 μl Plus reagent). After 6 hours, the media is changed to antibiotic-free DMEM with 10% fetal bovine serum. These methods use serum during cell culture, but serum-free methods are preferred.

Lentivirus can be purified as follows. Viral supernatants are harvested after 48 hours. Supernatants are first cleared of debris and filtered through a 0.45 μm low protein binding (PVDF) filter. They are then spun in an ultracentrifuge for 2 hours at 24,000 rpm. Viral

pellets are resuspended in 50 µl of DMEM overnight at 4° C. They are then aliquoted and immediately frozen at -80°C.

In another embodiment, minimal non-primate lentiviral vectors based on the equine infectious anemia virus (EIAV) are also contemplated. In another embodiment, RetinoStat®, an equine infectious anemia virus-based lentiviral gene therapy vector that expresses angiostatic proteins endostatin and angiostatin that is contemplated to be delivered via a subretinal injection. In another embodiment, use of self-inactivating lentiviral vectors are contemplated.

Any RNA of the systems, for example a guide RNA or a base editor-encoding mRNA, can be delivered in the form of RNA. Base editor-encoding mRNA can be generated using *in vitro* transcription. For example, nuclease mRNA can be synthesized using a PCR cassette containing the following elements: T7 promoter, optional kozak sequence (GCCACC), nuclease sequence, and 3' UTR such as a 3' UTR from beta globin-polyA tail. The cassette can be used for transcription by T7 polymerase. Guide polynucleotides (*e.g.*, gRNA) can also be transcribed using *in vitro* transcription from a cassette containing a T7 promoter, followed by the sequence "GG", and guide polynucleotide sequence.

To enhance expression and reduce possible toxicity, the base editor-coding sequence and/or the guide nucleic acid can be modified to include one or more modified nucleoside *e.g.* using pseudo-U or 5-Methyl-C.

The small packaging capacity of AAV vectors makes the delivery of a number of genes that exceed this size and/or the use of large physiological regulatory elements challenging. These challenges can be addressed, for example, by dividing the protein(s) to be delivered into two or more fragments, wherein the N-terminal fragment is fused to a split intein-N and the C-terminal fragment is fused to a split intein-C. These fragments are then packaged into two or more AAV vectors. As used herein, "intein" refers to a self-splicing protein intron (*e.g.*, peptide) that ligates flanking N-terminal and C-terminal exteins (*e.g.*, fragments to be joined). The use of certain inteins for joining heterologous protein fragments is described, for example, in Wood et al., J. Biol. Chem. 289(21); 14512-9 (2014). For example, when fused to separate protein fragments, the inteins IntN and IntC recognize each other, splice themselves out and simultaneously ligate the flanking N- and C-terminal exteins of the protein fragments to which they were fused, thereby reconstituting a full-length protein from the two protein fragments. Other suitable inteins will be apparent to a person of skill in the art.

A fragment of a fusion protein of the invention can vary in length. In some embodiments, a protein fragment ranges from 2 amino acids to about 1000 amino acids in length. In some embodiments, a protein fragment ranges from about 5 amino acids to about 500 amino acids in length. In some embodiments, a protein fragment ranges from about 20 amino acids to about 200 amino acids in length. In some embodiments, a protein fragment ranges from about 10 amino acids to about 100 amino acids in length. Suitable protein fragments of other lengths will be apparent to a person of skill in the art.

In one embodiment, dual AAV vectors are generated by splitting a large transgene expression cassette in two separate halves (5′ and 3′ ends, or head and tail), where each half of the cassette is packaged in a single AAV vector (of <5 kb). The re-assembly of the full-length transgene expression cassette is then achieved upon co-infection of the same cell by both dual AAV vectors followed by: (1) homologous recombination (HR) between 5′ and 3′ genomes (dual AAV overlapping vectors); (2) ITR-mediated tail-to-head concatemerization of 5′ and 3′ genomes (dual AAV trans-splicing vectors); or (3) a combination of these two mechanisms (dual AAV hybrid vectors). The use of dual AAV vectors *in vivo* results in the expression of full-length proteins. The use of the dual AAV vector platform represents an efficient and viable gene transfer strategy for transgenes of >4.7 kb in size.

Inteins

Inteins (intervening protein) are auto-processing domains found in a variety of diverse organisms, which carry out a process known as protein splicing. Protein splicing is a multi-step biochemical reaction comprised of both the cleavage and formation of peptide bonds. While the endogenous substrates of protein splicing are proteins found in intein-containing organisms, inteins can also be used to chemically manipulate virtually any polypeptide backbone.

In protein splicing, the intein excises itself out of a precursor polypeptide by cleaving two peptide bonds, thereby ligating the flanking extein (external protein) sequences via the formation of a new peptide bond. This rearrangement occurs post-translationally (or possibly co-translationally). Intein-mediated protein splicing occurs spontaneously, requiring only the folding of the intein domain.

About 5% of inteins are split inteins, which are transcribed and translated as two separate polypeptides, the N-intein and C-intein, each fused to one extein. Upon translation, the intein fragments spontaneously and non-covalently assemble into the canonical intein

structure to carry out protein splicing in trans. The mechanism of protein splicing entails a series of acyl-transfer reactions that result in the cleavage of two peptide bonds at the intein-extein junctions and the formation of a new peptide bond between the N- and C-exteins. This process is initiated by activation of the peptide bond joining the N-extein and the N-terminus of the intein. Virtually all inteins have a cysteine or serine at their N-terminus that attacks the carbonyl carbon of the C-terminal N-extein residue. This N to O/S acyl-shift is facilitated by a conserved threonine and histidine (referred to as the TXXH motif (SEQ ID NO: 17)), along with a commonly found aspartate, which results in the formation of a linear (thio)ester intermediate. Next, this intermediate is subject to trans-(thio)esterification by nucleophilic attack of the first C-extein residue (+1), which is a cysteine, serine, or threonine. The resulting branched (thio)ester intermediate is resolved through a unique transformation: cyclization of the highly conserved C-terminal asparagine of the intein. This process is facilitated by the histidine (found in a highly conserved HNF motif) and the penultimate histidine and may also involve the aspartate. This succinimide formation reaction excises the intein from the reactive complex and leaves behind the exteins attached through a non-peptidic linkage. This structure rapidly rearranges into a stable peptide bond in an intein-independent fashion. In some embodiments, the split intein is selected from Gp41.1, IMPDH.1, NrdJ.1 and Gp41.8 (Carvajal-Vallejos, Patricia *et al.* "Unprecedented rates and efficiencies revealed for new natural split inteins from metagenomic sources." J. Biol. Chem., vol. 287,34 (2012)).

Non-limiting examples of inteins include any intein or intein-pair known in the art, which include a synthetic intein based on the dnaE intein, the Cfa-N (e.g., split intein-N) and Cfa-C (e.g., split intein-C) intein pair, has been described (e.g., in Stevens et al., J Am Chem Soc. 2016 Feb. 24; 138(7):2162-5, incorporated herein by reference), and DnaE. Non-limitine examples of pairs of inteins that may be used in accordance with the present disclosure include: Cfa DnaE intein, Ssp GyrB intein, Ssp DnaX intein, Ter DnaE3 intein, Ter ThyX intein, Rma DnaB intein and Cne Prp8 intein (e.g., as described in U.S. Patent No. 8,394,604, incorporated herein by reference). Exemplary nucleotide and amino acid sequences of inteins are provided in the Sequence Listing at SEQ ID NOs: 482-489.

Intein-N and intein-C may be fused to the N-terminal portion of a split Cas9 and the C-terminal portion of the split Cas9, respectively, for the joining of the N-terminal portion of the split Cas9 and the C-terminal portion of the split Cas9. For example, in some embodiments, an intein-N is fused to the C-terminus of the N-terminal portion of the split Cas9, *i.e.*, to form a structure of N--[N-terminal portion of the split Cas9]-[intein-N]--C. In

some embodiments, an intein-C is fused to the N-terminus of the C-terminal portion of the split Cas9, *i.e.*, to form a structure of N-[intein-C]--[C-terminal portion of the split Cas9]-C. The mechanism of intein-mediated protein splicing for joining the proteins the inteins are fused to (*e.g.*, split Cas9) is known in the art, *e.g.*, as described in Shah *et al.*, Chem Sci. 2014; 5(1):446-461, incorporated herein by reference. Methods for designing and using inteins are known in the art and described, for example by WO2014004336, WO2017132580, US20150344549, and US20180127780, each of which is incorporated herein by reference in their entirety.

In some embodiments, a portion or fragment of a nuclease (*e.g.*, Cas9) is fused to an intein. The nuclease can be fused to the N-terminus or the C-terminus of the intein. In some embodiments, a portion or fragment of a fusion protein is fused to an intein and fused to an AAV capsid protein. The intein, nuclease and capsid protein can be fused together in any arrangement (*e.g.*, nuclease-intein-capsid, intein-nuclease-capsid, capsid-intein-nuclease, etc.). In some embodiments, an N-terminal fragment of a base editor (*e.g.*, ABE, CBE) is fused to a split intein-N and a C-terminal fragment is fused to a split intein-C. In some embodiments, an N-terminal fragment of a nucleic acid programmable DNA binding protein (napDNAbp) domain (*e.g.*, Cas9) is fused to a split intein-N and a C-terminal fragment is fused to a split intein-C. In some embodiments, an N-terminal fragment of a deaminase domain (*e.g.*, adenosine or cytidine deaminase) fused to a split intein-N and a C-terminal fragment is fused to a split intein-C.

These fragments are then packaged into two or more AAV vectors. In some embodiments, a polynucleotide encoding a base editor (*e.g.*, self-inactivating base editor) featuring an intein comprises an intron. In some embodiments, the N-terminus of an intein is fused to the C-terminus of a fusion protein and the C-terminus of the intein is fused to the N-terminus of an AAV capsid protein.

In one embodiment, inteins are utilized to join fragments or portions of a cytidine or adenosine base editor protein that is grafted onto an AAV capsid protein. The use of certain inteins for joining heterologous protein fragments is described, for example, in Wood et al., J. Biol. Chem. 289(21); 14512-9 (2014). For example, when fused to separate protein fragments, the inteins IntN and IntC recognize each other, splice themselves out and simultaneously ligate the flanking N- and C-terminal exteins of the protein fragments to which they were fused, thereby reconstituting a full-length protein from the two protein fragments. Other suitable inteins will be apparent to a person of skill in the art.

In some embodiments, an ABE was split into N- and C- terminal fragments at Ala, Ser, Thr, or Cys residues within selected regions of SpCas9. These regions correspond to loop regions identified by Cas9 crystal structure analysis.

The N-terminus of each fragment is fused to an intein-N and the C- terminus of each fragment is fused to an intein C at amino acid positions S303, T310, T313, S355, A456, S460, A463, T466, S469, T472, T474, C574, S577, A589, and S590, which are indicated in capital letters in the sequence below (called the "Cas9 reference sequence").

```
   1 mdkkysigld igtnsvgwav itdeykvpsk kfkvlgntdr hsikknliga llfdsgetae
  61 atrlkrtarr rytrrknric ylqeifsnem akvddsffhr leesflveed kkherhpifg
 121 nivdevayhe kyptiyhlrk klvdstdkad lrliylalah mikfrghfli egdlnpdnsd
 181 vdklfiqlvq tynqlfeenp inasgvdaka ilsarlsksr rlenliaqlp gekknglfgn
 241 lialslgltp nfksnfdlae daklqlskdt ydddldnlla qigdqyadlf laaknlsdai
 301 llSdilrvnT eiTkaplsas mikrydehhq dltllkalvr qqlpekykei ffdqSkngya
 361 gyidggasqe efykfikpil ekmdgteell vklnredllr kqrtfdngsi phqihlgelh
 421 ailrrqedfy pflkdnreki ekiltfripy yvgplArgnS rfAwmTrkSe eTiTpwnfee
 481 vvdkgasaqs fiermtnfdk nlpnekvlpk hsllyeyftv yneltkvkyv tegmrkpafl
 541 sgeqkkaivd llfktnrkvt vkqlkedyfk kieCfdSvei sgvedrfnAS lgtyhdllki
 601 ikdkdfldne enediledіv ltltlfedre mieerlktya hlfddkvmkq lkrrrytgwg
 661 rlsrklingi rdkqsgktil dflksdgfan rnfmqlihdd sltfkediqk aqvsgqgdsl
 721 hehianlags paikkgilqt vkvvdelvkv mgrhkpeniv iemarenqtt qkgqknsrer
 781 mkrieegike lgsqilkehp ventqlqnek lylyylqngr dmyvdqeldi nrlsdydvdh
 841 ivpqsflkdd sidnkvltrs dknrgksdnv pseevvkkmk nywrqllnak litqrkfdnl
 901 tkaergglse ldkagfikrq lvetrqitkh vaqildsrmn tkydendkli revkvitlks
 961 klvsdfrkdf qfykvreinn yhhahdayln avvgtalikk ypklesefvy gdykvydvrk
1021 miakseqeig katakyffys nimnffktei tlangeirkr plietngetg eivwdkgrdf
1081 atvrkvlsmp qvnivkktev qtggfskesi lpkrnsdkli arkkdwdpkk yggfdsptva
1141 ysvlvvakve kgkskklksv kellgitime rssfeknpid fleakgykev kkdliiklpk
1201 yslfelengr krmlasagel qkgnelalps kyvnflylas hyeklkgspe dneqkqlfve
1261 qhkhyldeii eqisefskrv iladanldkv lsaynkhrdk pireqaenii hlftltnlga
1321 paafkyfdtt idrkrytstk evldatlihq sitglyetri dlsqlggd
```

(SEQ ID NO:250).

Pharmaceutical Compositions

In some aspects, the present invention provides a pharmaceutical composition comprising any of the polynucleotides, vectors, cells, base editors (*e.g.*, self-inactivating base editor), base editor systems, guide polynucleotides, fusion proteins, or the fusion protein-guide polynucleotide complexes described herein

The pharmaceutical compositions of the present invention can be prepared in accordance with known techniques. See, *e.g.*, Remington, The Science And Practice of Pharmacy (21st ed. 2005). In general, the cell, or population thereof is admixed with a suitable carrier prior to administration or storage, and in some embodiments, the pharmaceutical composition further comprises a pharmaceutically acceptable carrier. Suitable pharmaceutically acceptable carriers generally comprise inert substances that aid in administering the pharmaceutical composition to a subject, aid in processing the pharmaceutical compositions into deliverable preparations, or aid in storing the pharmaceutical composition prior to administration. Pharmaceutically acceptable carriers can include agents that can stabilize, optimize or otherwise alter the form, consistency, viscosity, pH, pharmacokinetics, solubility of the formulation. Such agents include buffering agents, wetting agents, emulsifying agents, diluents, encapsulating agents, and skin penetration enhancers. For example, carriers can include, but are not limited to, saline, buffered saline, dextrose, arginine, sucrose, water, glycerol, ethanol, sorbitol, dextran, sodium carboxymethyl cellulose, and combinations thereof.

Some nonlimiting examples of materials which can serve as pharmaceutically-acceptable carriers include: (1) sugars, such as lactose, glucose and sucrose; (2) starches, such as corn starch and potato starch; (3) cellulose, and its derivatives, such as sodium carboxymethyl cellulose, methylcellulose, ethyl cellulose, microcrystalline cellulose and cellulose acetate; (4) powdered tragacanth; (5) malt; (6) gelatin; (7) lubricating agents, such as magnesium stearate, sodium lauryl sulfate and talc; (8) excipients, such as cocoa butter and suppository waxes; (9) oils, such as peanut oil, cottonseed oil, safflower oil, sesame oil, olive oil, corn oil and soybean oil; (10) glycols, such as propylene glycol; (11) polyols, such as glycerin, sorbitol, mannitol and polyethylene glycol (PEG); (12) esters, such as ethyl oleate and ethyl laurate; (13) agar; (14) buffering agents, such as magnesium hydroxide and aluminum hydroxide; (15) alginic acid; (16) pyrogen-free water; (17) isotonic saline; (18) Ringer's solution; (19) ethyl alcohol; (20) pH buffered solutions; (21) polyesters, polycarbonates and/or polyanhydrides; (22) bulking agents, such as polypeptides and amino acids (23) serum alcohols, such as ethanol; and (23) other non-toxic compatible substances employed in pharmaceutical formulations. Wetting agents, coloring agents, release agents, coating agents, sweetening agents, flavoring agents, perfuming agents, preservative and antioxidants can also be present in the formulation.

Pharmaceutical compositions can comprise one or more pH buffering compounds to maintain the pH of the formulation at a predetermined level that reflects physiological pH,

such as in the range of about 5.0 to about 8.0. The pH buffering compound used in the aqueous liquid formulation can be an amino acid or mixture of amino acids, such as histidine or a mixture of amino acids such as histidine and glycine. Alternatively, the pH buffering compound is preferably an agent which maintains the pH of the formulation at a predetermined level, such as in the range of about 5.0 to about 8.0, and which does not chelate calcium ions. Illustrative examples of such pH buffering compounds include, but are not limited to, imidazole and acetate ions. The pH buffering compound may be present in any amount suitable to maintain the pH of the formulation at a predetermined level.

Pharmaceutical compositions can also contain one or more osmotic modulating agents, *i.e.*, a compound that modulates the osmotic properties (*e.g.*, tonicity, osmolality, and/or osmotic pressure) of the formulation to a level that is acceptable to the blood stream and blood cells of recipient individuals. The osmotic modulating agent can be an agent that does not chelate calcium ions. The osmotic modulating agent can be any compound known or available to those skilled in the art that modulates the osmotic properties of the formulation. One skilled in the art may empirically determine the suitability of a given osmotic modulating agent for use in the inventive formulation. Illustrative examples of suitable types of osmotic modulating agents include, but are not limited to: salts, such as sodium chloride and sodium acetate; sugars, such as sucrose, dextrose, and mannitol; amino acids, such as glycine; and mixtures of one or more of these agents and/or types of agents. The osmotic modulating agent(s) may be present in any concentration sufficient to modulate the osmotic properties of the formulation.

In addition to a modified cell, or population thereof, and a carrier, the pharmaceutical compositions of the present invention can include at least one additional therapeutic agent useful in the treatment of disease. For example, some embodiments of the pharmaceutical composition described herein further comprises a chemotherapeutic agent. In some embodiments, the pharmaceutical composition further comprises a cytokine peptide or a nucleic acid sequence encoding a cytokine peptide. In some embodiments, the pharmaceutical compositions comprising the cell or population thereof can be administered separately from an additional therapeutic agent.

One consideration concerning the therapeutic use of genetically modified cells of the invention is the quantity of cells necessary to achieve an optimal or satisfactory effect. The quantity of cells to be administered may vary for the subject being treated. In one embodiment, between 10^4 to 10^{10}, between 10^5 to 10^9, or between 10^6 and 10^8 genetically modified cells of the invention are administered to a human subject. In some embodiments,

at least about 1 x 10^8, 2 x 10^8, 3 x 10^8, 4 x 10^8, and 5 x 10^8 genetically modified cells of the invention are administered to a human subject. Determining the precise effective dose may be based on factors for each individual subject, including their size, age, sex, weight, and condition. Dosages can be readily ascertained by those skilled in the art from this disclosure and the knowledge in the art.

The skilled artisan can readily determine the number of cells and amount of optional additives, vehicles, and/or carriers in compositions and to be administered in methods of the invention. Typically, additives (in addition to the cell(s)) are present in an amount of 0.001 to 50 % (weight) solution in phosphate buffered saline, and the active ingredient is present in the order of micrograms to milligrams, such as about 0.0001 to about 5 wt%, preferably about 0.0001 to about 1 wt%, still more preferably about 0.0001 to about 0.05 wt% or about 0.001 to about 20 wt%, preferably about 0.01 to about 10 wt%, and still more preferably about 0.05 to about 5 wt %. Of course, for any composition to be administered to an animal or human, and for any particular method of administration, it is preferred to determine therefore: toxicity, such as by determining the lethal dose (LD) and LD50 in a suitable animal model (*e.g.*, a rodent such as a mouse); and, the dosage of the composition(s), concentration of components therein, and the timing of administering the composition(s), which elicit a suitable response. Such determinations do not require undue experimentation from the knowledge of the skilled artisan, this disclosure and the documents cited herein. And, the time for sequential administrations can be ascertained without undue experimentation.

In some embodiments, the pharmaceutical composition is formulated for delivery to a subject. Suitable routes of administrating the pharmaceutical composition described herein include, without limitation: topical, subcutaneous, transdermal, intradermal, intralesional, intraarticular, intraperitoneal, intravesical, transmucosal, gingival, intradental, intracochlear, transtympanic, intraorgan, epidural, intrathecal, intramuscular, intravenous, intravascular, intraosseus, periocular, intratumoral, intracerebral, and intracerebroventricular administration.

In some embodiments, the pharmaceutical composition described herein is administered locally to a diseased site. In some embodiments, the pharmaceutical composition described herein is administered to a subject by injection, by means of a catheter, by means of a suppository, or by means of an implant, the implant being of a porous, non-porous, or gelatinous material, including a membrane, such as a sialastic membrane, or a fiber.

In other embodiments, the pharmaceutical composition described herein is delivered in a controlled release system. In one embodiment, a pump can be used (see, *e.g.*, Langer, 1990, Science 249: 1527-1533; Sefton, 1989, CRC Crit. Ref. Biomed. Eng. 14:201; Buchwald et al., 1980, Surgery 88:507; Saudek et al., 1989, N. Engl. J. Med. 321:574). In another embodiment, polymeric materials can be used. (See, *e.g.*, Medical Applications of Controlled Release (Langer and Wise eds., CRC Press, Boca Raton, Fla., 1974); Controlled Drug Bioavailability, Drug Product Design and Performance (Smolen and Ball eds., Wiley, New York, 1984); Ranger and Peppas, 1983, Macromol. Sci. Rev. Macromol. Chem. 23:61. See also Levy et al., 1985, Science 228: 190; During et al., 1989, Ann. Neurol. 25:351; Howard et al., 1989, J. Neurosurg. 71: 105.) Other controlled release systems are discussed, for example, in Langer, supra.

In some embodiments, the pharmaceutical composition is formulated in accordance with routine procedures as a composition adapted for intravenous or subcutaneous administration to a subject, *e.g.*, a human. In some embodiments, pharmaceutical composition for administration by injection are solutions in sterile isotonic use as solubilizing agent and a local anesthetic such as lignocaine to ease pain at the site of the injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water free concentrate in a hermetically sealed container such as an ampoule or sachette indicating the quantity of active agent. Where the pharmaceutical is to be administered by infusion, it can be dispensed with an infusion bottle containing sterile pharmaceutical grade water or saline. Where the pharmaceutical composition is administered by injection, an ampoule of sterile water for injection or saline can be provided so that the ingredients can be mixed prior to administration.

A pharmaceutical composition for systemic administration can be a liquid, *e.g.*, sterile saline, lactated Ringer's or Hank's solution. In addition, the pharmaceutical composition can be in solid forms and re-dissolved or suspended immediately prior to use. Lyophilized forms are also contemplated. The pharmaceutical composition can be contained within a lipid particle or vesicle, such as a liposome or microcrystal, which is also suitable for parenteral administration. The particles can be of any suitable structure, such as unilamellar or plurilamellar, so long as compositions are contained therein. Compounds can be entrapped in "stabilized plasmid-lipid particles" (SPLP) containing the fusogenic lipid dioleoylphosphatidylethanolamine (DOPE), low levels (5-10 mol%) of cationic lipid, and stabilized by a polyethyleneglycol (PEG) coating (Zhang Y. P. et al., Gene Ther. 1999, 6:

1438-47). Positively charged lipids such as N-[l-(2,3-dioleoyloxi)propyl]-N,N,N-trimethyl-amoniummethylsulfate, or "DOTAP," are particularly preferred for such particles and vesicles. The preparation of such lipid particles is well known. See, *e.g.*, U.S. Patent Nos. 4,880,635; 4,906,477; 4,911,928; 4,917,951; 4,920,016; and 4,921,757; each of which is incorporated herein by reference.

The pharmaceutical composition described herein can be administered or packaged as a unit dose, for example. The term "unit dose" when used in reference to a pharmaceutical composition of the present disclosure refers to physically discrete units suitable as unitary dosage for the subject, each unit containing a predetermined quantity of active material calculated to produce the desired therapeutic effect in association with the required diluent; *i.e.*, carrier, or vehicle.

Further, the pharmaceutical composition can be provided as a pharmaceutical kit comprising (a) a container containing a compound of the invention in lyophilized form and (b) a second container containing a pharmaceutically acceptable diluent (*e.g.*, sterile used for reconstitution or dilution of the lyophilized compound of the invention. Optionally associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration.

In another aspect, an article of manufacture containing materials useful for the treatment of the diseases described above is included. In some embodiments, the article of manufacture comprises a container and a label. Suitable containers include, for example, bottles, vials, syringes, and test tubes. The containers can be formed from a variety of materials such as glass or plastic. In some embodiments, the container holds a composition that is effective for treating a disease described herein and can have a sterile access port. For example, the container can be an intravenous solution bag or a vial having a stopper pierceable by a hypodermic injection needle. The active agent in the composition is a compound of the invention. In some embodiments, the label on or associated with the container indicates that the composition is used for treating the disease of choice. The article of manufacture can further comprise a second container comprising a pharmaceutically-acceptable buffer, such as phosphate-buffered saline, Ringer's solution, or dextrose solution. It can further include other materials desirable from a commercial and user standpoint, including other buffers, diluents, filters, needles, syringes, and package inserts with instructions for use.

In some embodiments, any of the fusion proteins, gRNAs, and/or complexes described herein are provided as part of a pharmaceutical composition. In some embodiments, the pharmaceutical composition comprises any of the fusion proteins provided herein. In some embodiments, the pharmaceutical composition comprises any of the complexes provided herein. In some embodiments, the pharmaceutical composition comprises a ribonucleoprotein complex comprising an RNA-guided nuclease (*e.g.*, Cas9) that forms a complex with a gRNA and a cationic lipid. In some embodiments pharmaceutical composition comprises a gRNA, a nucleic acid programmable DNA binding protein, a cationic lipid, and a pharmaceutically acceptable excipient. Pharmaceutical compositions can optionally comprise one or more additional therapeutically active substances.

In some embodiments, compositions provided herein are administered to a subject, for example, to a human subject, in order to effect a targeted genomic modification within the subject. In some embodiments, cells are obtained from the subject and contacted with any of the pharmaceutical compositions provided herein. In some embodiments, cells removed from a subject and contacted *ex vivo* with a pharmaceutical composition are re-introduced into the subject, optionally after the desired genomic modification has been effected or detected in the cells. Methods of delivering pharmaceutical compositions comprising nucleases are known, and are described, for example, in U.S. Patent Nos. 6,453,242; 6,503,717; 6,534,261; 6,599,692; 6,607,882; 6,689,558; 6,824,978; 6,933,113; 6,979,539; 7,013,219; and 7,163,824, the disclosures of all of which are incorporated by reference herein in their entireties. Although the descriptions of pharmaceutical compositions provided herein are principally directed to pharmaceutical compositions which are suitable for administration to humans, it will be understood by the skilled artisan that such compositions are generally suitable for administration to animals or organisms of all sorts, for example, for veterinary use.

Modification of pharmaceutical compositions suitable for administration to humans in order to render the compositions suitable for administration to various animals is well understood, and the ordinarily skilled veterinary pharmacologist can design and/or perform such modification with merely ordinary, if any, experimentation. Subjects to which administration of the pharmaceutical compositions is contemplated include, but are not limited to, humans and/or other primates; mammals, domesticated animals, pets, and commercially relevant mammals such as cattle, pigs, horses, sheep, cats, dogs, mice, and/or rats; and/or birds, including commercially relevant birds such as chickens, ducks, geese, and/or turkeys.

Formulations of the pharmaceutical compositions described herein can be prepared by any method known or hereafter developed in the art of pharmacology. In general, such preparatory methods include the step of bringing the active ingredient(s) into association with an excipient and/or one or more other accessory ingredients, and then, if necessary and/or desirable, shaping and/or packaging the product into a desired single- or multi-dose unit. Pharmaceutical formulations can additionally comprise a pharmaceutically acceptable excipient, which, as used herein, includes any and all solvents, dispersion media, diluents, or other liquid vehicles, dispersion or suspension aids, surface active agents, isotonic agents, thickening or emulsifying agents, preservatives, solid binders, lubricants and the like, as suited to the particular dosage form desired. Remington's The Science and Practice of Pharmacy, 21st Edition, A. R. Gennaro (Lippincott, Williams & Wilkins, Baltimore, MD, 2006; incorporated in its entirety herein by reference) discloses various excipients used in formulating pharmaceutical compositions and known techniques for the preparation thereof. See also PCT application PCT/US2010/055131 (Publication number WO2011/053982 A8, filed Nov. 2, 2010), incorporated in its entirety herein by reference, for additional suitable methods, reagents, excipients and solvents for producing pharmaceutical compositions comprising a nuclease.

Except insofar as any conventional excipient medium is incompatible with a substance or its derivatives, such as by producing any undesirable biological effect or otherwise interacting in a deleterious manner with any other component(s) of the pharmaceutical composition, its use is contemplated to be within the scope of this disclosure.

The compositions, as described above, can be administered in effective amounts. The effective amount will depend upon the mode of administration, the particular condition being treated, and the desired outcome. It may also depend upon the stage of the condition, the age and physical condition of the subject, the nature of concurrent therapy, if any, and like factors well-known to the medical practitioner. For therapeutic applications, it is that amount sufficient to achieve a medically desirable result.

In some embodiments, compositions in accordance with the present disclosure can be used for treatment of any of a variety of diseases, disorders, and/or conditions.

Methods of Treatment

Some aspects of the present invention provide methods of treating a subject in need, the method comprising administering to a subject in need an effective therapeutic amount of a

pharmaceutical composition as described herein. More specifically, the methods of treatment include administering to a subject in need thereof one or more pharmaceutical compositions comprising one or more cells having at least one edited gene. In other embodiments, the methods of the invention comprise expressing or introducing into a cell a base editor polypeptide (*e.g.*, self-inactivating base editor) and one or more guide RNAs capable of targeting a nucleic acid molecule encoding at least one polypeptide

In one embodiment, a subject is administered at least 0.1×10^5 cells, at least 0.5×10^5 cells, at least 1×10^5 cells, at least 5×10^5 cells, at least 1×10^6 cells, at least 0.5×10^7 cells, at least 1×10^7 cells, at least 0.5×10^8 cells, at least 1×10^8 cells, at least 0.5×10^9 cells, at least 1×10^9 cells, at least 2×10^9 cells, at least 3×10^9 cells, at least 4×10^9 cells, at least 5×10^9 cells, or at least 1×10^{10} cells. In particular embodiments, about 1×10^7 cells to about 1×10^9 cells, about 2×10^7 cells to about 0.9×10^9 cells, about 3×10^7 cells to about 0.8×10^9 cells, about 4×10^7 cells to about 0.7×10^9 cells, about 5×10^7 cells to about 0.6×10^9 cells, or about 5×10^7 cells to about 0.5×10^9 cells are administered to the subject.

In one embodiment, a subject is administered at least 0.1×10^4 cells/kg of bodyweight, at least 0.5×10^4 cells/kg of bodyweight, at least 1×10^4 cells/kg of bodyweight, at least 5×10^4 cells/kg of bodyweight, at least 1×10^5 cells/kg of bodyweight, at least 0.5×10^6 cells/kg of bodyweight, at least 1×10^6 cells/kg of bodyweight, at least 0.5×10^7 cells/kg of bodyweight, at least 1×10^7 cells/kg of bodyweight, at least 0.5×10^8 cells/kg of bodyweight, at least 1×10^8 cells/kg of bodyweight, at least 2×10^8 cells/kg of bodyweight, at least 3×10^8 cells/kg of bodyweight, at least 4×10^8 cells/kg of bodyweight, at least 5×10^8 cells/kg of bodyweight, or at least 1×10^9 cells/kg of bodyweight. In particular embodiments, about 1×10^6 cells/kg of bodyweight to about 1×10^8 cells/kg of bodyweight, about 2×10^6 cells/kg of bodyweight to about 0.9×10^8 cells/kg of bodyweight, about 3×10^6 cells/kg of bodyweight to about 0.8×10^8 cells/kg of bodyweight, about 4×10^6 cells/kg of bodyweight to about 0.7×10^8 cells/kg of bodyweight, about 5×10^6 cells/kg of bodyweight to about 0.6×10^8 cells/kg of bodyweight, or about 5×10^6 cells/kg of bodyweight to about 0.5×10^8 cells/kg of bodyweight are administered to the subject.

One of ordinary skill in the art would recognize that multiple administrations of the pharmaceutical compositions contemplated in particular embodiments may be required to affect the desired therapy. For example, a composition may be administered to the subject 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more times over a span of 1 week, 2 weeks, 3 weeks, 1 month, 2 months, 3 months, 4 months, 5 months, 6 months, 1 year, 2 years, 5, years, 10 years, or more. In any of such methods, the methods may comprise administering to the subject an

effective amount of an edited cell or a base editor system or polynucleotide encoding such system. In any of such methods, the methods may comprise administering one or more doses of an effective amount of the edited cells per day. In any of such methods, the methods may comprise administering two or more doses of an effective amount of the edited cells per day. In any of such methods, the methods may comprise administering three or more doses of an effective amount of the edited cells per day. In any of such methods, the methods may comprise administering one or more doses of an effective amount of the edited cells per week. In any of such methods, the methods may comprise administering two or more doses of an effective amount of the edited cells per week. In any of such methods, the methods may comprise administering three or more doses of an effective amount of the edited cells per week. In any of such methods, the methods may comprise administering one or more doses of an effective amount of the edited cells per month. In any of such methods, the methods may comprise administering two or more doses of an effective amount of the edited cells per month. In any of such methods, the methods may comprise administering three or more doses of an effective amount of the edited cells per month.

Administration of the pharmaceutical compositions contemplated herein may be carried out using conventional techniques including, but not limited to, infusion, transfusion, or parenterally. In some embodiments, parenteral administration includes infusing or injecting intravascularly, intravenously, intramuscularly, intraarterially, intrathecally, intratumorally, intradermally, intraperitoneally, transtracheally, subcutaneously, subcuticularly, intraarticularly, subcapsularly, subarachnoidly and intrasternally.

In some embodiments, a composition described herein (*e.g.*, edited cell, base editor system) is administered in a dosage that is about 0.5-30 mg per kilogram body weight of the human subject. In another embodiment, the amount of the composition administered is about 0.5-20 mg per kilogram body weight of the human subject. In another embodiment, the amount of the composition administered is about 0.5-10 mg per kilogram body weight of the human subject. In another embodiment, the amount of the composition administered is about 0.04 mg, about 0.08 mg, about 0.16 mg, about 0.32 mg, about 0.64 mg, about 1.25 mg, about 1.28 mg, about 1.92 mg, about 2.5 mg, about 3.56 mg, about 3.75 mg, about 5.0 mg, about 7.12 mg, about 7.5 mg, about 10 mg, about 14.24 mg, about 15 mg, about 20 mg, or about 30 mg per kilogram body weight of the human subject. In another embodiment, the amount of the compo composition und administered is about 1.92 mg, about 3.75 mg, about 7.5 mg, about 15.0 mg, or about 30.0 mg per kilogram body weight of the human subject and the composition is administered two times a week. In another embodiment, the amount of the

composition administered is about 1.28 mg, about 2.56 mg, about 5.0 mg, about 10 mg, or about 20 mg per kilogram body weight of the human subject and the composition is administered two times a week. In another embodiment, the amount of the composition administered is about 1.92 mg, about 3.75 mg, about 7.5 mg, about 15.0 mg, or about 30.0 mg per kilogram body weight of the human subject and the composition is administered once a week. In another embodiment, the amount of the composition administered is about 1.28 mg, about 2.56 mg, about 5.0 mg, about 10 mg, or about 20 mg per kilogram body weight of the human subject and the composition is administered once a week. In another embodiment, the amount of the composition administered is about 1.92 mg, about 3.75 mg, about 7.5 mg, about 15.0 mg, or about 30.0 mg per kilogram body weight of the human subject and the composition is administered once a day three, five or seven times in a seven day period. In another embodiment, the composition is administered intravenously once a day, seven times in a seven day period. In another embodiment, the amount of the composition administered is about 1.28 mg, about 2.56 mg, about 5.0 mg, about 10 mg, or about 20 mg per kilogram body weight of the human subject and the composition is administered once a day three, five or seven times in a seven day period. In another embodiment, the composition is administered intravenously once a day, seven times in a seven day period.

In some embodiments, the composition is administered over a period of 0.25 h, 0.5 h, 1 h, 2 h, 3 h, 4 h, 5 h, 6 h, 7 h, 8 h, 9 h, 10 h, 11 h, or 12 h. In another embodiment, the composition is administered over a period of 0.25-2 h. In another embodiment, the composition is gradually administered over a period of 1 h. In another embodiment, the composition is gradually administered over a period of 2 h.

Kits

The invention provides kits featuring a self-inactivating a base editor. In some embodiments, the kit includes a polynucleotide comprising a heterologous intron encoding a self-inactivating base editor. In some embodiments, the intron is present in a polynucleotide that encodes a nucleic acid programmable DNA binding protein (napDNAbp) or a fragment thereof. In some embodiments, the intron is present in a polynucleotide that encodes a deaminase. In some embodiments, the kit further includes one or more guide polynucleotides (*e.g.*, a guide polynucleotide targeting a genomic sequence, a guide polynucleotide targeting a heterologous intron present in a polynucleotide encoding a self-inactivating base editor). In some embodiments, the kit includes a base editor system comprising a self-inactivating base editor comprising a nucleic acid programmable DNA binding protein (napDNAbp) and a

deaminase, and one or more guide polynucleotides. In some embodiments, the kit includes a vector comprising a polynucleotide encoding a self-inactivating base editor, wherein an open reading frame of the polynucleotide comprises a heterologous intron. In some embodiments, the kit includes a vector comprising a polynucleotide encoding one or more guide polynucleotides. In some embodiments, the kit includes a cell including any of the polynucleotides, base editors, base editor systems, cells or vectors as provided herein.

The kits may further comprise written instructions for using or inactivating a base editor described herein (*e.g.*, self-inactivating base editor). In other embodiments, the instructions include at least one of the following: precautions; warnings; clinical studies; and/or references. The instructions may be printed directly on the container (when present), or as a label applied to the container, or as a separate sheet, pamphlet, card, or folder supplied in or with the container. In a further embodiment, a kit can comprise instructions in the form of a label or separate insert (package insert) for suitable operational parameters. In yet another embodiment, the kit can comprise one or more containers with appropriate positive and negative controls or control samples, to be used as standard(s) for detection, calibration, or normalization. The kit can further comprise a second container comprising a pharmaceutically-acceptable buffer, such as (sterile) phosphate-buffered saline, Ringer's solution, or dextrose solution. It can further include other materials desirable from a commercial and user standpoint, including other buffers, diluents, filters, needles, syringes, and package inserts with instructions for use.

The practice of the present invention employs, unless otherwise indicated, conventional techniques of molecular biology (including recombinant techniques), microbiology, cell biology, biochemistry and immunology, which are well within the purview of the skilled artisan. Such techniques are explained fully in the literature, such as, "Molecular Cloning: A Laboratory Manual", second edition (Sambrook, 1989); "Oligonucleotide Synthesis" (Gait, 1984); "Animal Cell Culture" (Freshney, 1987); "Methods in Enzymology" "Handbook of Experimental Immunology" (Weir, 1996); "Gene Transfer Vectors for Mammalian Cells" (Miller and Calos, 1987); "Current Protocols in Molecular Biology" (Ausubel, 1987); "PCR: The Polymerase Chain Reaction", (Mullis, 1994); "Current Protocols in Immunology" (Coligan, 1991). These techniques are applicable to the production of the polynucleotides and polypeptides of the invention, and, as such, may be considered in making and practicing the invention. Particularly useful techniques for particular embodiments will be discussed in the sections that follow.

The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the assay, screening, and therapeutic methods of the invention, and are not intended to limit the scope of what the inventors regard as their invention.

EXAMPLES

EXAMPLE 1: Self-inactivation by catalytic inactivation of the deaminase domain

Traditional adenine base editors (ABE) utilize an evolved TadA enzyme to deaminate adenine DNA bases, resulting in 2'-deoxy inosine products, which are recognized by the cell as 2'-deoxyguanosine during DNA repair (Gaudelli, N. M., *et al.* (2017). *Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage*. Nature, 551(7681), 464-471; Gaudelli, N. M., *et al.* (2020). *Directed evolution of adenine base editors with increased activity and therapeutic application*. Nat Biotechnol, 38(7), 892-900; Richter, M. F., *et al.* (2020). *Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity*. Nature Biotechnology, 38(7), 883-891). TadA contains four catalytic residues responsible for deamination activity: His57 (H57), Glu59 (E59), Cys87 (C87), or Cys90 (C90) (Kim, J., *et al.* (2006). *Structural and kinetic characterization of Escherichia coli TadA, the wobble-specific tRNA deaminase*. Biochemistry, 45(20), 6407-6416).

A proof-of-concept assay was established to test whether 1) these sites might be leveraged for self-inactivation of ABE after delivery of its encoding DNA to cells, and 2) co-delivery of DNA encoding two guide RNAs and ABE, may efficiently and simultaneously edit a genomic target site of interest and the episomal DNA encoding ABE to convert a TadA catalytic residue to an inactive amino acid (FIG. 1A). The assay utilized plasmid lipofections of HEK293T cells in which the DNA vectors encoded ABE, a guide RNA targeting one of the four TadA catalytic residues (His57 (H57), Glu59 (E59), Cys87 (C87), or Cys90 (C90)), and a second guide RNA targeting a site within the cellular genome. Cells were incubated for 5 days after lipofection, and Illumina next-generation sequencing was performed on each target amplicon generated from cell lysates, which included both cellular genomic DNA and recovered plasmid DNA. To confirm that the TadA residues were catalytically inactive if edited, the corresponding mutations were pre-installed in control plasmids (providing amino acid substitutions H57R, E59G, C87R, or C90R) in the TadA subunit of ABE7.10-m, and their base editing activity was assessed on the cellular genomic target site alone. In each case, the pre-installed mutation produced virtually no editing (<0.1% A-to-G conversion) at

the site of interest, confirming that these TadA mutations were inactivating (FIG. 1B). In samples transfected with functional ABE7.10-m, editing at the genomic site and editing on the recovered base editor plasmid were both assessed. Only two of the target sites in TadA (H57R and E59G) were appreciably edited by their corresponding self-inactivating guide RNA, while editing rates on the genomic target were similar to that of a control containing a scrambled self-inactivating guide RNA. This revealed that while rates of self-editing at the catalytic sites were relatively low (<5% A-to-G conversion), targeting the DNA encoding the base editor ABE was possible, and on-target editing efficiency within the host genome was not negatively impacted by self-inactivation of the editor.

To determine whether the rate of self-inactivation could be improved, the codon usage within TadA was optimized to create an optimized spacer sequence for editing to occur. Given the available NGG PAMs within TadA, the target adenosine of the E59 codon is positioned in what is predicted to be a more favorable position within the editing window of ABE7.10 as well as more recently described ABE8 variants (position A7 for E59, versus position A9 for H57) (FIG. 1C). Thus, the E59 site was optimized. All possible synonymous codon usages of TadA in the target region were bioinformatically analyzed and ranked by prioritizing high relative synonymous codon usage (RSCU) scores and prioritizing low in silico off-target prediction profiles for each corresponding guide RNA. Seventeen candidates were then tested in the context of ABE8.5-m, which had separately demonstrated higher editing efficiency on the cellular genomic target of interest compared to ABE7.10-m. Using the plasmid-based assay, self-inactivation rates of editing were varied and low overall (<5%) (FIG. 1D), while editing at the cellular genomic site ranged from approximately 30–50% across variants. Four of the new variants, two of the first-generation variants, and a first-generation variant featuring a scrambled-guide RNA—all of which were in the context of ABE8.5-m—were subsequently packaged into AAV2 and transduced into ARPE-19 cells. Resulting edits within the cellular genome and on the AAV transgene were assessed at multiple timepoints up to 5-weeks. Although the cellular genomic locus was edited in approximately 70–76% of NGS reads across all tested variants at 5 weeks (FIG. 1E), the abundance of self-inactivation edits was low (3.54%) in the recovered editor DNA at the same time point for the most optimized variant (FIG. 1F). Compared to the unoptimized first-generation variant in the same assay, however, self-inactivation rates of the new variants were improved up to 10.4-fold.

While the apparent rates of self inactivation of ABE DNA were generally low across plasmid-based and viral-based in vitro assays, it was suspected that it was possible that measuring

self inactivation on the recovered episomal DNA could be underreporting the functional inactivation rate. For example, it has been reported that only a fraction of transduced AAV2 is trafficked to the cell nucleus in vitro (Xiao, P.-J., Li, C., *et al.* (2012). *Quantitative 3D Tracing of Gene-delivery Viral Vectors in Human Cells and Animal Tissues*. Molecular Therapy, 20(2), 317-328). If some portion of AAV or LNP particles in the assays remained encapsulated within the cytoplasm, they would presumably not be accessible to the base editor, nor would they be transcribed to make additional base editor mRNA and protein. Because the mRNA within the cells is more indicative of the protein that is being expressed at any given time, performed targeted amplicon sequencing was performed for both the DNA as well as cDNA that was prepared from the mRNA to examine any differences in the measured amount of self-inactivation in a new experiment. Five AAV2-packaged editor variants and guide RNAs were transduced into ARPE-19 cells. At the genomic target site, editing was robust (~80% A-to-G editing) across all five variants tested after 2-weeks post transduction (FIG. 1G). Self-inactivation rates were again similar to previous studies with other variants (<3%) when measured on the recovered episomal DNA, but the targeted amplicon sequencing of the cDNA prepared from mRNA showed that the proportion of edited ABE transcripts approached an average of 48.5% across the five variants. These results suggested that the proportion of ABE DNA that was successfully delivered to the nucleus and actively transcribed was efficiently self-inactivated—approaching nearly 50% inactivation after two weeks—and that the simultaneous edit at the genomic target remained highly efficient. These results also suggested that a similar approach could be undertaken for the self-inactivation of CBE where, for example, the APOBEC active site residues responsible for coordinating zinc (Cys or His) could be targeted for C-to-T base conversion to generate Tyr in either case.

EXAMPLE 2: Self-inactivation by mutation of the base editor start codon

In another approach, the start codon of ABE was targeted for self-inactivation by base editing the first nucleotide of the editor start codon, resulting in a Met1Val mutation (ATG>GTG). Alternatively, base editing at the second nucleotide of the start codon by targeting the complementary DNA strand is also possible, resulting in a Met1Thr mutation (ATG>ACG). To first determine if these mutations would effectively prevent translation of the base editor, the desired mutations were preinstalled into the ABE8.5-m expression plasmid. When Met1 is mutated, there is potential for production of a truncated editor if the downstream Met12 of TadA functions as an alternative start codon. In an effort to circumvent this issue, the codon usage of His8 was adjusted, such that an out-of-frame ATG sequence is present upstream from Met12, and base editing at the site of interest within the

cellular genome was measured to determine the ability of these mutations to prevent editor expression (FIG. 2A). The Met1Thr mutation more effectively prevented editing at the site of interest than Met1Val, and the addition of the out-of-frame start codon further reduced editing by half (FIG. 2B).

In order to perform the Met1Thr base edit on ABE8.5-m DNA containing a nickase variant of the *Streptococcus pyogenes* Cas9, an NGG PAM sequence was required within TadA at a location that would place the target base for self-inactivation inside the editing window of the protospacer. To make this accommodation, variants of the editor with TadA mutations Phe8Trp, Phe8Val, or Phe8Ala were prepared to generate three different NGG PAM sequences that could be utilized (FIG. 2C) for base editing in Met1. Each variant was capable of editing the cellular genomic site with equal efficiency compared to an unaltered control ABE8.5-m, suggesting that these mutations of Phe8 in TadA are well tolerated. Self-inactivation rates measured by target amplicon sequencing of the recovered episomal editor DNA were variable and low (<3% base conversion) for each variant, and the simultaneous edits at the genomic target across all new variants were as efficient as the control ABE8.5-m that lacked an inactivating guide RNA (FIG. 2E). These results demonstrated that it was also possible to self-inactivate the base editor by converting the start codon to a non-Met amino acid via a single A-to-G base edit without sacrificing editing efficiency at a desired site within the cellular genome. These results also suggested that a similar approach could be undertaken for the self-inactivation of CBE where the Met start codon could be converted to Ile to prevent or inhibit the initiation of translation.

EXAMPLE 3: Incorporation of introns into polynucleotides encoding base editors for customized, self-inactivation

To achieve the maximum utility of ABE self-inactivation, the ideal approach should be generalizable across any chosen variant of ABE, which may vary by PAM requirement, editing window, and relative activity on a given target site. Strategies for self-inactivation of ABE by editing within TadA active site residues, the start codon, or any other specific residue within the editor are each constrained by the local amino acid sequence of the target site and any permissive mutations that are tolerated by the enzyme. To develop a solution to these limitations, the utility of incorporating small introns into the DNA encoding ABE was explored (FIG. 3A). Introns generally feature a consensus splice donor sequence, GT, at their 5'-end and a consensus splice acceptor sequence, AG, at their 3'-end. Each of these sites can be targeted by ABE to interrupt splicing by converting the consensus sequences to GC

and GG, respectively. When either splice site within the editor is disrupted, the intron sequence may be fully or partially retained, which can result in premature termination or a frameshift in the base editor sequence. The key advance to this approach is that it provides for flexibility in design of the self-inactivation target sequence, as the intron can be customized and inserted within any conducive part of the editor that provides an ideal editing position, protospacer, and PAM for promoting efficient base editing as dictated by a chosen base editor variant.

Intron-based inactivation of CBE is also possible through disruption of the canonical splice donor or acceptor site via C-to-T editing of the non-coding DNA strand, respectively converting GT to AT or converting AG to AA. Furthermore, the intron could alternatively be incorporated into other subunits of the editor such as Cas9 or linker regions, which may be more advantageous for inactivating circularly permuted base editor variants or other configurations of the editor (Huang, T. P., *et al.* (2019). *Circularly permuted and PAM-modified Cas9 variants broaden the targeting scope of base editors.* Nat Biotechnol, 37(6), 626-631; Nguyen Tran, M. T., *et al.* (2020) *Engineering domain-inlaid SaCas9 adenine base editors with reduced RNA off-targets and increased on-target DNA editing.* Nat Commun, 11(1), 4871).

The evolved TadA component of ABE was originally derived from bacteria and therefore lacks introns in its native context. To determine if inclusion of an intron was detrimental to ABE expression, editing rates were measured at a genomic target site catalyzed by intron-containing ABE variants and compared to a control ABE that lacked an intron. No self-inactivating guide RNAs were included in the experiment. To keep the total size of the editor as small as possible, only small introns ($\leq$ ~150 bp) were tested. A set of nine human introns and one chimeric intron, which was derived from human beta globin and mouse IgG genes, were inserted directly after codon 87 of TadA. Additionally, a subset of introns was also inserted directly after codon 62, or after the second nucleotide of codon 23 in TadA to determine if the location of the intron impacted editing rates. Each of the intron locations were chosen for their relative proximity to an NGG PAM sequence within the TadA coding sequence that would situate the required intron edit within the known base editing window of the editor variant being examined. The editing efficiency of these intron-containing variants at a genomic target site in HEK293T cells was dependent on both the intron sequence and the location of insertion within TadA (Fig 3B), suggesting that some configurations were more efficiently spliced than others. Five different introns at TadA residue 87 (NF1, PAX2, EEF1A1, Chimera, and SLC50A1) maintained high editing efficiency on the genomic DNA.

Additionally, introns Chimera and ABCB11 permitted relatively high levels of editing when inserted at residues 62 or 23 of TadA. ABE variants containing the remaining four introns (BRSK2, PLXNB3, TMPRSS6, and IL32) either produced no editing at the site of interest or resulted in a significant decrease in the editing rate compared to the control. The activity of seven more introns at TadA residue 87 were tested and were all equally well tolerated (FIG. 3C). These data indicated that a custom intron sequence could be inserted into the TadA subunit of the ABE gene without sacrificing editing activity at a genomic site.

In this system, introns were selected that feature the consensus splice donor and acceptor sequence found in eukaryotes. However, these dinucleotide sequences are not completely conserved across all eukaryotic introns, and their sensitivity to mutations may be variable. Therefore, the desired splice acceptor site or donor site edits were pre-installed within the TadA introns, and base editing activity was assessed at a genomic site to determine their corresponding levels of inactivation compared to ABE8.5-m lacking an intron. Introns NF1, PAX2, EEF1A1, ANTXRL, PKHD1L1, PADI1, HMCN2, ENPEP-gecko, and HMCN2-salmon at TadA residue 87 were fully disrupted by a preinstalled splice acceptor mutation (FIG. 3D). These nine intron configurations were also among the most efficient ABE variants in the absence of the mutation, as shown in the previous experiments, suggesting that these variants functioned in a binary fashion depending on the presence or absence of a single A-to-G base edit at the splice acceptor site. The NF1 and PAX2 introns were also inactivated by the splice donor mutation, but EEF1A1 was not. The Chimera intron was the only variant that was completely unperturbed by mutations at the splice donor or acceptor site. Additional configurations were also tested with and without preinstalled splice acceptor or donor mutations, and these further demonstrated that the NF1 intron placed at TadA residues 18, 59, or 129 was fully inactivated by a splice acceptor mutation, and EEF1A1 was inactivated by the splice acceptor mutation when placed at TadA residues 18 or 59 (FIG. 3E). These data demonstrated that splicing of an unnatural intron within TadA could be fully disrupted by an A-to-G substitution at the site of splicing, and that the edit is capable of ablating base editing activity on DNA. Furthermore, these data demonstrated that the placement of the intron with the editor could affect the sensitivity to splice site mutations and the overall splicing performance of the intron.

To determine which intron variants could be efficiently base edited at their respective splice sites, HEK293T cells were lipofected with plasmids encoding an ABE variant, a matching self-inactivating guide RNA, and a guide RNA targeting a genomic site. Targeting the splice donor sites of introns NF1 and PAX2 at TadA residue 87 maintained high rates of

editing at the genomic site, and the degree of self-inactivation editing within the intron donor sites was as high as ~6% in the recovered plasmid DNA (FIG. 3F). Alternatively, the amount of self-inactivation edit at the splice acceptor site of NF1 at TadA residue 87 was 30%, while also maintaining efficient editing at the genomic site (FIG 3G).

Additional configurations of introns that had performed well at TadA residue 87 (NF1, PAX2, and EEF1A1) were evaluated. This set was inserted after codons 129, 59, and 18 of TadA and the corresponding, preinstalled splice-acceptor mutations were also prepared. The editing efficiency of the NF1 variants lacking the preinstalled mutation was varied depending on where the intron was inserted (FIG. 3H). PAX2 and EEF1A1 maintained high genomic editing in all configurations, but these introns were not consistently inactivated across all locations within TadA when a mutation was preinstalled.

The ability of each of these variants to be targeted by a self-inactivating guide RNA was examined. The rate of self-inactivation of NF1 at TadA residue 59 was similar to that of NF1 at TadA residue 87 (FIG. 3I). In addition, intron sequences ANTXRL, PKHD1L1, PADI1 and ENPEP-gecko were tested at TadA residue 87 for guide-dependent self-inactivation at the splice acceptor site, since this set of introns had been previously validated in this configuration as being fully inactivated by the mutation. These were compared to NF1, PAX2, and EEF1A1 in the same experiment, which revealed that NF1 was the most efficiently targeted by its self-inactivating guide RNA, followed by PAX2 and ANTXRL (FIG. 3J). Each of the remaining introns displayed similar, low levels of guide RNA-dependent self-inactivation. Notably, base editing activity on the genomic site was not hampered by the presence of a self-inactivating guide RNA in any of these examples compared to an ABE8.5-m variant lacking an intron or self-inactivating guide RNA. These data demonstrated that introns could be strategically placed within TadA to provide an amenable base-editing target site, which can be utilized to shut-off base editing activity over time in a guide RNA-dependent manner. These results also demonstrated that the rate of self-inactivation was slow enough to simultaneously allow for efficient base editing at a desired site within the host genomic DNA that is directed by an additional guide RNA.

To determine if the rates of self-inactivation by sequencing the recovered episomal DNA was underreporting the functional level of self-inactivation in intron-based studies, lipofections were reperformed with plasmids expressing ABE8.5-m containing the NF1 or PAX2 introns at TadA residue 87, a matching self-inactivating guide RNA targeting the splice acceptor site, and a second guide RNA targeting a genomic site. By target amplicon sequencing, the amount of editing observed in the genomic DNA from cells treated with

either intron variant was similar to that of a control ABE8.5-m editor lacking an intron or self-targeting guide RNA. The amount of self-inactivation edits measured from the recovered episomal DNA were 5.38% and 4.16%, respectively, for NF1 and PAX2. In order to determine the functional inactivation rate from the RNA of the cells, RNAseq analysis was performed on the total mRNA in order to avoid possible PCR bias that may occur due to the size differences between the spliced and unspliced ABE transcripts. In stark contrast to the perceived inactivation rates measured on the plasmids, >84% of NF1 transcripts contained an edited, unspliced intron, and only 3.7% of transcripts were functionally spliced (FIG. 3L). The majority of the remaining ABE transcripts (11.25%) contained unedited introns that were retained within the transcript, which should lead to premature termination during translation, and a small fraction of transcripts (0.66%) were alternatively spliced. Similarly, >79% of PAX2 transcripts contained an edited, unspliced intron, and ~16.5% of transcripts were functionally spliced (FIG. 3M). These data suggest that at the conclusion of this 5-day experiment, the majority of ABE DNA that was being actively transcribed contained an edited intron sequence that prevented splicing from occurring, and it could therefore be reasonably assumed that base editor expression would have been reduced as well.

Finally, the NF1 intron, positioned at TadA residue 87, was tested by AAV delivery. The editor and guide RNAs were packaged into AAV2, which was used to transduce ARPE-19 cells *in vitro*, and editing rates were assessed at 2-weeks post-transduction. By target amplicon sequencing, the genomic DNA editing was robust at the desired target site (~76% A-to-G), but editing rates in the recovered AAV genomes were low (~5.4% A-to-G) at the intronic splice acceptor site (FIG. 3N). In contrast, RNA-seq analysis on the total mRNA showed much higher self-inactivation efficiency, where ~88.5% of transcripts contained the edited, retained intron. This data demonstrated that base editors delivered by AAV could be self-inactivated by alteration of splicing on a timescale that also allows for an efficient genomic edit to be performed by a second guide RNA.

EXAMPLE 4: Evaluation of efficiencies of simultaneous self-inactivation and base editing of a target genomic site

Experiments were undertaken to evaluate the efficiencies for simultaneous self-inactivation and target (i.e., "desired") genomic site base editing using base editor systems containing an ABE8.5-m and different combinations of guide RNA's. One of the guide RNAs was used to edit a target genomic site and the other was used for self-inactivation of the ABE8.5-m.

An experiment was undertaken to evaluate efficiencies for self-inactivation and target-site base editing using two different strategies for self-inactivation. The strategies included 1) self-inactivation through alteration of the TadA catalytic site of ABE8.5-m through the introduction of an E59G alteration, and 2) self-inactivation through alteration of a splice-acceptor site of an NF1 or PAX2 intron inserted within a polynucleotide encoding the ABE8.5-m. The introns were inserted within a polynucleotide encoding the ABE8.5-m following codon 87 of the TadA deaminase of the ABE8.5-m. The E59G self-inactivating alteration was introduced using the guide RNA v1, v122, v224, v139, v110, or v113 (sequences are provided in **Table 1B**). The self-inactivating alteration to the splice-acceptor site of NF1 was introduced using the guide g235 (see **Table 1B**), and the self-inactivating alteration to the splice acceptor site of PAX2 was introduced using the guide g239 (see **Table 1B**). Throughout this example, the guide used to edit the target genomic site was g756 (see **Table 1B**). ARPE-19 cells were infected using AAV2 viral particles containing polynucleotides encoding the ABE8.5-m, with or without an NF1 or PAX2 intron insertion, in combination with a guide RNAs to facilitate base editing of the target genomic site and a self-inactivating base-edit to a portion of the polynucleotide encoding the TadA deaminase domain of the ABE8.5-m, as indicated in **FIGs. 4A-4C**. Percent (%) desired (i.e., "target") base conversion at the genomic site and for self-inactivation of ABE8.5-m were measured using DNA next-generation sequencing and/or RNA-seq. Measurements were taken at 1, 3, and 5 weeks post-transduction. As negative controls, cells were also transduced using a sequence-scrambled ("scrmbl") guide. As shown in **FIGs. 4A-4C**, the ABE8.5-m base editor systems were capable of simultaneously self-inactivating and editing the target genomic site.

Next, a titration experiment was undertaken to determine the effect of AAV2 dose on base editing of the target genomic site and self-inactivation of ABE8.5-m (see **FIGs. 5A and 5B**). The ABE8.5-m was self-inactivated either through alteration of a splice-acceptor site of an NF1 intron inserted within the ABE8-5-m, as described above, or through introduction of the E59G alteration, as described above, using the guide RNA v110 (see **Table 1B**). ARPE-19 cells were infected using AAV2 viral particles containing polynucleotides encoding ABE8.5-m with or without an NF1 intron insertion in combination with guide RNAs to facilitate base editing of the target genomic site and a self-inactivating base-edit to a portion of the polynucleotide encoding the TadA deaminase domain of the ABE8.5-m, as indicated in **FIGs. 5A and 5B**. Cells were transduced using the AAV2 particles at doses of 89k vg/cell (high), 17 k vg/cell (medium), or 9k vg/cell (low). It was determined that both efficiency of base editing at the target genomic site increased with dosage while efficiency of self-

inactivation remained the same across the dosages evaluated, as determined through next-generation DNA sequencing measurements taken at 3, 7, and 14 weeks post-transduction. It was also observed that base-editing efficiencies for both self-inactivation and altering the target genomic site increased over time.

Next, an experiment was undertaken to determine the efficiencies for simultaneous target genomic site base editing and self-inactivation using the two self-inactivation strategies described above (see **FIGs. 6A and 6B**). The E59G self-inactivating alteration was introduced using the guide RNA v110 (see **Table 1B**). The self-inactivating alteration to the splice-acceptor site of NF1 was introduced using the guide g235, and the self-inactivating alteration to the splice acceptor site of PAX2 was introduced using the guide g239. The guide used to edit the target genomic site was g756 (see **Table 1B**). ARPE-19 cells were infected using AAV2 viral particles containing polynucleotides encoding ABE8.5-m, with or without an NF1 or PAX2 intron insertion, in combination with guide RNAs to facilitate base editing of the target genomic site and a self-inactivating base-edit to a portion of the polynucleotide encoding the TadA deaminase domain of the ABE8.5-m, as indicated in **FIGs. 6A and 6B.** Percent (%) desired (i.e., "target") base conversion at the genomic site and for self-inactivation of ABE8.5-m were measured using DNA next-generation sequencing and RNA-seq, respectively. Measurements were taken at 4, 7, and 14 weeks post-transduction. As shown in **FIGs. 6A and 6B**, the ABE8.5-m base editor systems were capable of simultaneously self-inactivating and editing the target genomic site.

An experiment was undertaken to determine the efficiencies for simultaneous target genomic site base editing and self-inactivation using the two self-inactivation strategies described above (see **FIGs. 7A and 7B**), where the cells were transduced using plasmids (i.e., lipofection). The E59G self-inactivating alteration was introduced using the guide RNA v110 (see **Table 1B**). The self-inactivating alteration to the splice-acceptor site of NF1 was introduced using the guide g235. The guide used to edit the target genomic site was g756 (see **Table 1A**). HEK293T cells were contacted with lipid nano particles containing polynucleotides encoding ABE8.5-m, with or without an NF1 intron insertion, in combination with a guide RNA to facilitate base editing of the target genomic site and a self-inactivating base-edit to a portion of the polynucleotide encoding the TadA deaminase domain of the ABE8.5-m, as indicated in **FIGs. 7A and 7B.** Percent (%) desired (i.e., "target") base conversion at the genomic site and for self-inactivation of ABE8.5-m were measured using DNA next-generation sequencing and RNA-seq, respectively. As negative controls, cells were also transduced using sequence-scrambled ("scrmbl") guides.

Measurements were taken at 2 and 7 weeks post-transduction. As shown in **FIGs. 7A and 7B**, the ABE8.5-m base editor systems were capable of simultaneously self-inactivating and editing the target genomic site, and efficiency of self-inactivation and editing of the target genomic site both increased with time.

Next, an experiment was undertaken to determine the efficiencies for simultaneous target genomic site base editing and self-inactivation using the two self-inactivation strategies described above (see **FIGs. 8A and 8B**), where cells were transduced using AAV8 viral particles. The E59G self-inactivating alteration was introduced using the guide RNA v110 (see **Table 1B**). The self-inactivating alteration to the splice-acceptor site of NF1 was introduced using the guide g235. The guide used to edit the target genomic site was g756 (see **Table 1B**). ARPE-19 cells were infected using AAV8 viral particles containing polynucleotides encoding ABE8.5-m, with or without an NF1 intron insertion, in combination with a guide RNA to facilitate base editing of the target genomic site and a self-inactivating base-edit to a portion of the polynucleotide encoding the TadA deaminase domain of the ABE8.5-m, as indicated in **FIGs. 8A and 8B.** Percent (%) desired (i.e., "target") base conversion at the genomic site and for self-inactivation of ABE8.5-m were measured using DNA next-generation sequencing and/or RNA-seq. As negative controls, cells were also transduced using sequence-scrambled ("scrmbl") guides. Measurements were taken at 4 (**FIG. 8A**) and 4 (**FIG. 8B**) weeks post-transduction. As shown in **FIGs. 8A and 8B**, the ABE8.5-m base editor systems were capable of simultaneously self-inactivating and editing the target genomic site.

The following materials and methods were employed in the above examples.

General methods

The constructs used in Examples 1-3 were generated through USER enzyme cloning, restriction enzyme digest and T4 DNA ligation, Gibson assembly, or synthesized by Genewiz. All PCR reactions used either Phusion U DNA Polymerase Green MultiPlex PCR Master Mix (Thermo Fisher) or Q5 Hot Start HiFi 2x Master Mix (New England Biolabs). The introns were ordered as gene fragments from Integrated DNA Technologies (IDT). All experiments, including those performed by plasmid lipofection and those performed by AAV transductions, utilized a split base editor that was recombined using a Cfa(GEP) split intein fusion, where the editor was split between at the amino acid residues corresponding Asn309 and Thr310 of Cas9, and residue 310 was mutated to Thr310Cys. The split editor was

encoded by two separate plasmids or AAV vectors which required co-transfection or co-transduction, respectively. Each ABE fragment was expressed from a CMV promoter. One or more guide RNAs were encoded in tandem on the vector that also encoded the N-terminal split of the base editor, and the guide RNAs were expressed using U6 promoters. Guide RNAs were inserted into digested recipient plasmids by ligation of annealed, complementary oligos ordered from IDT. Guide oligos were annealed in IDT duplex buffer by heating sample to 95°C for 3 minutes and cooling to 20°C at a rate of 0.1°C/sec. Plasmid DNA was prepared using QIAGEN Plasmid *Plus* Kits from 35 ml of either Mach1 (Thermo Fisher) or NEB Stable cell culture (New England Biolabs) grown in LB media containing 100 µg/mL carbenicillin. In all experiments that describe targeting a genomic target site, the desired edit corresponded to A-to-G conversion of the ABCA4 c.5882G>A point mutation that had been lentivirally integrated into all cell lines used in the study. Editing activity resulting from lipofections or transductions of cells were measured from technical replicates.

AAV production and titration

rAAV vectors were produced using transient triple transfection of suspension cultures of HEK293T cells with plasmids containing the (a) AAV Rep and Cap genes, (b) transgene flanked by inverted terminal repeat (ITR) sequences, and (c) adenovirus genes (E4, E2a and VA) necessary for AAV replication (helper plasmid). 72 hrs after transfection, cells were lysed and unpackaged DNA was removed by adding Triton X-100, $MgCl_2$ and Turbonuclease at a final concentration of 0.25% (v/v), 2 mM, and 10 U/ml, respectively. The lysis was performed at 37°C for 2-5 hrs in a shaker incubator. The cell lysate was filtered through a clarification depth filter, followed by a 0.2-micron filtration. Clarified lysate was loaded onto an affinity chromatography column using a Cytiva AKTA Pure chromatography system. Captured AAV was eluted using an elution buffer at pH 2.5-3. Eluate was captured, and the pH was promptly neutralized. Full and empty AAV particles were separated by cesium chloride density gradient ultracentrifugation. Bands containing full AAV particles were collected, and the titer was determined by qPCR. The full particle samples were then diluted in CsCl stock solution to a final target concentration and dialyzed into formulation buffer (10 mM Na_2HPO_4, 2 mM KH_2PO_4, 2.7 mM KCl, 192 mM NaCl, 0.001% Pluronic F-68; pH 7.4) using appropriately sized dialysis cassettes (MWCO 100 kDa). Dialysate was filtered using low protein binding 0.1-micron syringe filters and then aliquoted and stored at –80°C. Final AAV titers were determined using droplet digital PCR (ddPCR) from vialed material. Other

quality attributes such as endotoxin level, aggregation status, osmolality, and pH were also determined.

Generation of cell lines containing an integrated target site of interest using a lentiviral insertion

The lentiviral production plasmid was generated by restriction cloning using a 5' HpaI and 3' ApaI flanked g-block (ordered from IDT) that encoded a human ABCA4 gene fragment comprising the sequence starting 72-bp before exon 42 through 123-bp after exon 42 and which also included a point mutation corresponding to ABCA4 c.5882G>A. The g-block was ligated into an HpaI/ApaI digested pLenti6.4 R4R2 V5-DEST vector (Thermo Fisher Scientific), and the final plasmid was used for small-scale lentiviral production at Vigene Biosciences. Lentivirus was used to transduce HEK293T cells [CRL-3216, American Type Cell Culture Collection (ATCC)] and ARPE-19 cells (CRL-2302, ATCC) at a multiplicity of infection (MOI) of 0.3–10 IFU/cell. Cells that were stably integrated were selected by growing on base media supplemented with 10 μg/mL blasticidin, and once established, the cells were maintained on base media supplemented with 5 μg/mL blasticidin. The average number of lentiviral integrations per cell within each population of cells pooled by MOI were assessed by target amplicon sequencing using primers oBTx361 and oBTx362 to simultaneously amplify the virally integrated ABCA4 fragment and the endogenous ABCA4 locus. The number of viral integrations in each sample was estimated by multiplying the number of NGS reads containing the ABCA4 c.5882G>A mutation by 2 divided by the number of wild type reads (integration number = mutant reads x 2 / WT reads). Cell line pools containing an average of two or fewer integrations per cell were used for all base editing experiments.

Cell culture of HEK293T

Cells were cultured at 37°C with 5% CO_2. Lentivirally integrated cells were maintained in Dulbecco's modified Eagles medium plus Glutamax (10566-016, Thermo Fisher Scientific) with 10% (v/v) fetal bovine serum (A31606-02, Thermo Fisher Scientific) and 5 μg/ml blasticidin S HCl (A1113903, Thermo Fisher Scientific) following the culture method provided by ATCC.

Cell culture of ARPE-19

Cells were cultured at 37°C with 5% CO_2. Lentivirally integrated cells were maintained in Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12 (DMEM/F-12) (11320033, Thermo Fisher Scientific) with 10% (v/v) fetal bovine serum (FBS) (A31606-02, Thermo Fisher Scientific) and 5 µg/ml blasticidin S HCl (A1113903, Thermo Fisher Scientific) following culture method provided by ATCC.

Plasmid transfection

HEK293T cells were seeded in Corning® CellBIND® 48-well Multiple Well Plates (3338, Corning) at a density of 35,000 cells per well in Dulbecco's modified Eagles medium plus Glutamax with 10% (v/v) FBS without blasticidin. Cells were transfected ~24 hours after seeding. Complementary plasmid pairs containing the split base editor and up to two guide RNAs were combined at a 1:1 molar ratio for a total of 1000 ng, followed by the addition of 1.5 µL Lipofectamine 2000 (Thermo Fisher Scientific) and Opti-MEM reduced serum medium (Thermo Fisher Scientific) for a total volume of 25 µl. The reagent mixtures were added to the wells following the manufacturer's instructions. Media was replaced every 48 hours over a 5-day (120 hour) period before cell lysis.

AAV transduction

ARPE-19 cells were seeded in Corning® CellBIND® 48-well Multiple Well Plates at a density of 23,000 cells per well in DMEM/F-12 media and 10% FBS without blasticidin. At approximately 24 hours after seeding, media was removed, and each of the two AAVs containing the split base editor was diluted to 50,000 viral genomes per virus per cell in a total volume of 125 µl per well of DMEM/F-12 lacking FBS. Cells were incubated with virus at 37°C with 5% CO_2 for 3 hours and 125 µl of DMEM/F-12 with 20% FBS was added to each sample for a final concentration of 10% FBS. Media was replaced with DMEM/F-12 and 10% FBS every 48 hours over a 14-day period before cell lysis for standard AAV transduction experiments. For time course experiments, replicates were lysed every 7 days over a 5-week period.

Extraction of genomic DNA and editor DNA

Media was removed and cells were washed twice with 100 µl of 1x PBS (Thermo Fisher) before adding 75 µl of cell lysis buffer (10 mM Tris·HCl (pH 8.0) + 0.05% SDS + 100 µg/mL Proteinase K). Wells were scraped with pipet tips (using multichannel) and

immediately transferred to a 96 well plate. The 96 well plate was incubated at 55°C for 1 hour followed by heat inactivation at 95°C for 20 minutes. Samples were then stored at - 20°C.

Alternatively, an additional wash step including Turbo DNAse (Thermo Fisher) was tested prior to lysis in an attempt to mitigate the potential for the presence of residual extracellular plasmid DNA impacting the self-inactivation analysis. The addition or absence of Turbo DNase yielded the same results by NGS analysis, so Turbo DNAse was excluded in subsequent experiments.

Reverse Transcription of mRNA to cDNA for target amplicon sequencing

Total RNA was isolated using the MagMAX™ mirVana™ Total RNA Isolation Kit (A27828, Thermo Fisher Scientific) following the manufacturer's protocol. Total RNA was reverse transcribed to cDNA using the RevertAid RT Reverse Transcription Kit (K1691, Thermo Fisher Scientific), following the manufacturer's protocol except that the random hexamer primer was excluded from the reaction, and an anchored oligo dT primer was used instead (5'- TTTTTTTTTTTTTTTTTTTTNN (SEQ ID NO: 481)). The reverse transcription reaction was incubated at 25°C for 10 min, 37°C for 60 min, and 95°C for 5 min. The crude cDNA was used directly in downstream applications.

Target amplicon sequencing DNA and cDNA samples

Cell lysate (2 µL) or cDNA was added to a 25 µL PCR reaction containing Q5 Hot Start HiFi 2x Master Mix and 0.5 µM of each primer containing 5' Illumina adapter overhangs. Each sample was amplified in two separate reactions with two unique primer pairs: one pair flanking the cellular genomic site of interest (oBTx360 and oBTx368), and another pair flanking the desired edit site of the editor DNA sequence. The primers oBTx360 and oBTx368 selectively amplify the lentivirally integrated genomic target site by priming to exon 42 of ABCA4 and a V5 tag flanking the ABCA4 fragment. PCR reactions were carried out as follows: 95°C for 2 min, 30 cycles of (95°C for 15 s, 65°C for 20 s, and 72°C for 20 s), and a final 72°C extension for 2 min. Following amplification, 2 µL of the crude PCR products containing the amplified site of interest were barcoded using 0.5 µM of each unique Illumina barcoding primer pairs and Q5 Hot Start High-Fidelity 2X Master Mix in a total volume of 25 µL. The reactions were carried out as follows: 98°C for 2 min, 10 cycles of (98°C for 20 s, 60°C for 30 s, and 72°C for 30 s), and a final 72°C extension for 2 min. Equal volumes of barcoded PCR products were then pooled and cleaned up using SPRISelect

paramagnetic beads (Beckman Coulter) using a 0.6X bead/sample ratio. Eluted DNA concentration was quantified with a Qubit 4 (Thermo Fisher Scientific) and was sequenced with an Illumina MiSeq instrument according to the manufacturer's protocol.

RNA-seq of total mRNA

Cells were cultured and transfected as previously described above. At the conclusion of the experiments, media was removed, and the cells were detached and dissociated with 50 µL TrypLE™ Express Enzyme (1X), phenol red (12605036, Thermo Fisher Scientific). The cells were then pelleted and washed once with complete media and then split into two pellets: one to be used for targeted sequencing of the DNA as described above, and one to be used for RNA-seq. Total RNA was isolated using the MagMAX™ mirVana™ Total RNA Isolation Kit (A27828, Thermo Fisher Scientific) following the manufacturer's protocol. Next, 200 ng of total RNA was processed using NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (E7760L, New England Biolabs) according to the manufacturer's instructions (protocol for use with NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB E7490)). Final cDNA products were quantified on a TapeStation (Agilent), normalized to 4nM, pooled by equal volume and then sequenced with an Illumina NextSeq 550 instrument following the manufacturer's protocol.

Amplicon sequencing analysis details

1. FASTQ files were generated from base call files (BCF) created by the MiSeq instrument using Illumina blc2fastq (v2.20.0.422) with the following parameters.

```
bcl2fastq \
    --ignore-missing-bcls \
    --ignore-missing-filter \
    --ignore-missing-positions \
    --ignore-missing-controls \
    --auto-set-to-zero-barcode-mismatches \
    --find-adapters-with-sliding-window \
    --adapter-stringency 0.9 \
    --mask-short-adapter-reads 35 \
    --minimum-trimmed-read-length 35 \
```

2. The FASTQ files created were then processed using trimmomatic (v0.39) with parameters set up to clip Illumina TruSeq adapters, exclude reads shorter than 20 bases, trim

the remaining 3' end of reads if the average base quality (Phred score) in a 4-bp sliding window dropped below 15, trim any bases with quality scores of 3 or lower at the end of reads, and trim the 4 randomized bases introduced from the round 1 PCR primers.

The following command was used to execute trimmomatic:

```
trimmomatic SE -phred33 $input_fastq $output_fastq \
        ILLUMINACLIP:illumine_adapters.fa:2:30:10 \
        LEADING:3 TRAILING:3 \
        SLIDINGWINDOW:4:15 \
        MINLEN:20 \
        HEADCROP:4
```

3. Trimmed reads were aligned to expected amplicon sequences using bowtie2 (v2.35) in end-to-end mode with the --very-sensitive flag specified. The SAM files created by bowtie2 were converted to BAM files, sorted, and indexed using samtools (v1.9).

4. The BAM files created in step (3) were processed using the bam-readcounts tool (github.com/genome/bam-readcount) to generate plain text files summarizing the number of non-reference bases, deletions and insertions at each position in the alignment. The minimum base quality (Phred score) for counting a non-reference base was set to 29 in order to exclude low confidence base calls from statistics about editing rates. Editing rates for each position in the target site were calculated as the fraction of non-reference bases of a given type (*e.g.*, G) to the total number of bases passing the base quality threshold at a given position in the alignment.

Total mRNA sequencing analysis details

Lane level FASTQ files were separately aligned to a custom-built genome including the human genome (Gencode GRCh38v31 primary assembly) and base editor construct using STAR (v2.7.2a) with parameters set to specify the ReadGroup and output both a genome aligned BAM file and a transcriptome aligned BAM file. Lane level genome alignments for each sample created in step were merged, sorted by coordinate, and duplicate marked using Picard (v2.20.5). BAM files were then loaded into IGV for visualization and quantitation of editing rates.

Other Embodiments

From the foregoing description, it will be apparent that variations and modifications may be made to the invention described herein to adopt it to various usages and conditions. Such embodiments are also within the scope of the following claims.

The recitation of a listing of elements in any definition of a variable herein includes definitions of that variable as any single element or combination (or subcombination) of listed elements. The recitation of an embodiment herein includes that embodiment as any single embodiment or in combination with any other embodiments or portions thereof.

All patents and publications mentioned in this specification are herein incorporated by reference to the same extent as if each independent patent and publication was specifically and individually indicated to be incorporated by reference.

CLAIMS

What is claimed is:

1. A polynucleotide encoding a deaminase domain or a nucleic acid programmable DNA binding protein (napDNAbp) domain or fragment thereof, the polynucleotide comprising an intron, wherein the intron is inserted in an open reading frame encoding the deaminase or a napDNAbp or fragment thereof.

2. A polynucleotide encoding a deaminase domain or a nucleic acid programmable DNA binding protein (napDNAbp) domain open reading frame comprising an intron, the intron comprising an alteration at a splice acceptor or splice donor site, wherein the alteration reduces or eliminates splicing of base editor mRNA, thereby reducing or eliminating expression of a base editor polypeptide.

3. A polynucleotide encoding a base editor polypeptide or fragment thereof, the polynucleotide comprising an intron, wherein the intron is inserted in an open reading frame encoding the base editor polypeptide or fragment thereof.

4. The polynucleotide of claim 3, wherein the base editor has high editing efficiency in genomic DNA.

5. A polynucleotide comprising a base editor open reading frame comprising an intron, the intron comprising an alteration at a splice acceptor or splice donor site, wherein the alteration reduces or eliminates splicing of base editor mRNA, thereby reducing or eliminating expression of a base editor polypeptide.

6. The polynucleotide of any one of claims 3-5, wherein the base editor comprises a nucleic acid programmable DNA binding protein (napDNAbp) domain or a deaminase domain.

7. A polynucleotide encoding a base editor comprising a nucleic acid programmable DNA binding protein (napDNAbp) domain or a deaminase domain, the polynucleotide comprising an intron, wherein the intron is inserted in an open reading frame encoding the napDNAbp domain or the deaminase domain.

8. A polynucleotide encoding a base editor comprising a nucleic acid programmable DNA binding protein (napDNAbp) domain, and a deaminase domain, or a fragment thereof, the polynucleotide comprising a base editor open reading frame comprising an intron, the intron comprising an alteration at a splice acceptor or splice donor site, wherein the alteration reduces splicing of the base editor mRNA.

9. The polynucleotide of any one of claims 1, 2 or 6-8, wherein the deaminase domain is a cytidine deaminase domain or an adenosine deaminase domain.

10. The polynucleotide of any one of claims 1, 2 or 6-9, wherein the napDNAbp domain is a Cas domain selected from the group consisting of a Cas9, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, and Cas12j/CasΦ domain.

11. The polynucleotide of any one of claims 1-10, wherein the intron is derived from a sequence selected from the group consisting of NF1, PAX2, EEF1A1, HBB, IGHG1, SLC50A1, ABCB11, BRSK2, PLXNB3, TMPRSS6, IL32, ANTXRL, PKHD1L1, PADI1, KRT6C, and HMCN2.

12. The polynucleotide of claim 11, wherein the intron is derived from NF1.

13. The polynucleotide of claim 11, wherein the intron is derived from PAX2.

14. The polynucleotide of claim 11, wherein the intron is derived from EEF1A1.

15. The polynucleotide of claim 11, wherein the intron is derived from HBB.

16. The polynucleotide of claim 11, wherein the intron is derived from IGHG1.

17. The polynucleotide of claim 11, wherein the intron is derived from SLC50A1.

18. The polynucleotide of claim 11, wherein the intron is derived from ABCB11.

19. The polynucleotide of claim 11, wherein the intron is derived from BRSK2.

20. The polynucleotide of claim 11, wherein the intron is derived from PLXNB3.

21. The polynucleotide of claim 11, wherein the intron is derived from TMPRSS6.

22. The polynucleotide of claim 11, wherein the intron is derived from IL32.

23. The polynucleotide of claim 11, wherein the intron is derived from PKHD1L1.

24. The polynucleotide of claim 11, wherein the intron is derived from PADI1.

25. The polynucleotide of claim 11, wherein the intron is derived from KRT6C.

26. The polynucleotide of claim 11, wherein the intron is derived from HMCN2.

27. The polynucleotide of any one of claims 1-26, wherein the intron has at least about 85% nucleic acid sequence identity to an intron naturally present in a mammalian gene.

28. The polynucleotide of any one of claims 1-26, wherein the intron has at least about 85% nucleic acid sequence identity to an intron naturally present in a non-mammalian gene.

29. The polynucleotide of any one of claims 1-10, wherein the intron is a synthetic intron.

30. The polynucleotide of any one of claims 1-26, wherein the intron comprises a sequence that has at least about 85% nucleic acid sequence identity to one of the following:

a) GTGAGATCAAATGAAAGTTTCATATAGAAATACAAAACCTAGAGAACTGGCATGT AAGAGAAGCAAAAATTACTTCAGCAAGGCCATGTTAGTAAATTTGCATCTGTTTGTCCACAT TAG (SEQ ID NO: 226);

b) GTAGGTGACAATGCTGCAGCTGCCTAATCTAGGTGGGGGGAACTAAATTGTGGGT GAGCTGCTGAATGGTCTGTAGTCTGAGGCTGGGGTGGGGGGAGACACAACGTCCCCTCCCTG CAAACCACTGCTATTCTGTCCCTCTCTCTCCTTAG (SEQ ID NO: 227);

c) GTAAGTGGCTTTCAAGACCATTGTTAAAAAGCTCTGGGAATGGCGATTTCATGCT TACATAAATTGGCATGCTTGTGTTTCAG (SEQ ID NO: 228);

d) GTAAGTATCAAGGTTACAAGACAGGTTTAAGGAGACCAATAGAAACTGGGCTTGT CTAGACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTACTGACATCCACTTT GCCTTTCTCTCCACAG (SEQ ID NO: 229);

e) GTAAGCACAACTGGGATGGGGTGACAGGGGTGCAAGATTGAAAACTGGCTCCTCT CCTCATAGCAGTTCTTGTGATTTCAG (SEQ ID NO: 230);

f) GTAAGAAATGTTATTTTTCAGTAAGTGATTTAGTTATTTTTCCTTTTTTTCTCATTA AAATTTCTCTAACATCTCCCTCTTCATGTTTTAG (SEQ ID NO: 231);

g) GTGAGACCCTAGCCCCCTCAACCCTGCCCTGGCCTCTCCCCAAACCTGCCCCCCC ACGCTGACCCCCACACCCGGCCGCCCGCAG (SEQ ID NO: 232);

h) GTGGGTGTCAGAGGCATCGGGGCTGCGGGGTAGGGGGCTGCCCCACCCCTAACGA AGTCTGCTCCTCCAG (SEQ ID NO: 233);

i) GCAGGGAAGTCCTGCTTCCGTGCCCCACCGGTGCTCAGCTGAGGCTCCCTTGAAAA TGCGAGGCTGTTTCCAACTTTGGTCTGTTTCCCTGGCAG (SEQ ID NO: 234);

j) GTGGGGAGTTGGGGTCCCCGAAGGTGAGGACCCTCTGGGGATGAGGGTGCTTCTCT GAGACACTTTCTTTTCCTCACACCTGTTCCTCGCCAGCAG (SEQ ID NO: 235);

k) GTATAGACCCCTTGATCTCCTAACCCTAACCCTAACCCTAACCCTAACCTACAAA ATCTTAGAGCATCAGTGGGAGCATCTCACTGTCCAGGCTCAATATTTCTTCATTTTCTTGCA G (SEQ ID NO: 236);

l) GTAATTATGATTAAAGATGGTGATTGTTTATTTTCTTTTATGATTGTCCTTAGTAT TATGTAACCTGCAAATTCTATTGCAG (SEQ ID NO: 237);

m) GTGAGTGACACAAGGTGTTGTCTGGGGAGTGGGGAAGGGGGATGGAAGTGAATCC TGTTGGTGGGGTGGAGAAAGGGCGATCTCAAGAGGGCCACTCTCTCCAG (SEQ ID NO: 238);

n) GTAAGCATCTCCACCATCCTTCTGTTTACTCTGATGGGGTCTGCAAAGGGGAGAT GATGTATAGGGTTGGGTATCTCTGTAAATGTCAGATGTGAAGTTGATCTTATGACCTTCTGT TCTGCAG (SEQ ID NO: 239);

o) GTGAGGGTCTCCCAGGCTGGGCAGGGGGAGGGGGCTGCTGCCTTGATTGCGTCCC AGGACACAGCCCTCCTCCAGCCTGCCCTCGCCTTGCTCATCCCCTCCCCATCTCAGCCCCAC CCCCACTAACTCTCTCTCTGCTCTGACTCAG (SEQ ID NO: 240);

p) GTAATGATTGATTGCAATGTATGATTACAATAATCTCAGTATAAGTTCAGTAATA ATAACCTTCCACTGCTGTCCTCTGTGTGCACCCAG (SEQ ID NO: 241); or

q) GTAAATATATACAACAGTTTTTCATTTAAATAAGTGCACGGCACAAATAAGAAAA ATATGTCAAAAATGTAACCAATAGTTTTTTTCAAATTTAG (SEQ ID NO: 242).

31. The polynucleotide of any one of claims 1-26, wherein the intron comprises a nucleic acid sequence from one of the following:

a) GTGAGATCAAATGAAAGTTTCATATAGAAATACAAAACCTAGAGAACTGGCATGT AAGAGAAGCAAAAATTACTTCAGCAAGGCCATGTTAGTAAATTTGCATCTGTTTGTCCACAT TAG (SEQ ID NO: 226);

b) GTAGGTGACAATGCTGCAGCTGCCTAATCTAGGTGGGGGGAACTAAATTGTGGGT GAGCTGCTGAATGGTCTGTAGTCTGAGGCTGGGGTGGGGGGAGACACAACGTCCCCTCCCTG CAAACCACTGCTATTCTGTCCCTCTCTCTCCTTAG (SEQ ID NO: 227);

c) GTAAGTGGCTTTCAAGACCATTGTTAAAAAGCTCTGGGAATGGCGATTTCATGCT TACATAAATTGGCATGCTTGTGTTTCAG (SEQ ID NO: 228);

d) GTAAGTATCAAGGTTACAAGACAGGTTTAAGGAGACCAATAGAAACTGGGCTTGT CTAGACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTACTGACATCCACTTT GCCTTTCTCTCCACAG (SEQ ID NO: 229);

e) GTAAGCACAACTGGGATGGGGTGACAGGGGTGCAAGATTGAAAACTGGCTCCTCT CCTCATAGCAGTTCTTGTGATTTCAG (SEQ ID NO: 230);

f) GTAAGAAATGTTATTTTTCAGTAAGTGATTTAGTTATTTTTCCTTTTTTTCTCATTA AAATTTCTCTAACATCTCCCTCTTCATGTTTTAG (SEQ ID NO: 231);

g) GTGAGACCCTAGCCCCCTCAACCCTGCCCTGGCCTCTCCCCAAACCTGCCCCCCC ACGCTGACCCCCACACCCGGCCGCCCGCAG (SEQ ID NO: 232);

h) GTGGGTGTCAGAGGCATCGGGGCTGCGGGGTAGGGGGCTGCCCCACCCCTAACGA AGTCTGCTCCTCCAG (SEQ ID NO: 233);

i) GCAGGGAAGTCCTGCTTCCGTGCCCCACCGGTGCTCAGCTGAGGCTCCCTTGAAAA TGCGAGGCTGTTTCCAACTTTGGTCTGTTTCCCTGGCAG (SEQ ID NO: 234);

j) GTGGGGAGTTGGGGTCCCCGAAGGTGAGGACCCTCTGGGGATGAGGGTGCTTCTCT GAGACACTTTCTTTTCCTCACACCTGTTCCTCGCCAGCAG (SEQ ID NO: 235);

k) GTATAGACCCCTTGATCTCCTAACCCTAACCCTAACCCTAACCCTAACCTACAAA ATCTTAGAGCATCAGTGGGAGCATCTCACTGTCCAGGCTCAATATTTCTTCATTTTCTTGCA G (SEQ ID NO: 236);

l) GTAATTATGATTAAAGATGGTGATTGTTTATTTTCTTTTATGATTGTCCTTAGTAT TATGTAACCTGCAAATTCTATTGCAG (SEQ ID NO: 237);

m) GTGAGTGACACAAGGTGTTGTCTGGGGAGTGGGGAAGGGGGATGGAAGTGAATCC TGTTGGTGGGGTGGAGAAAGGGCGATCTCAAGAGGGCCACTCTCTCCAG (SEQ ID NO: 238);

n) GTAAGCATCTCCACCATCCTTCTGTTTACTCTGATGGGGTCTGCAAAGGGGAGAT GATGTATAGGGTTGGGTATCTCTGTAAATGTCAGATGTGAAGTTGATCTTATGACCTTCTGT TCTGCAG (SEQ ID NO: 239);

o) GTGAGGGTCTCCCAGGCTGGGCAGGGGGAGGGGGCTGCTGCCTTGATTGCGTCCC AGGACACAGCCCTCCTCCAGCCTGCCCTCGCCTTGCTCATCCCCTCCCCATCTCAGCCCCAC CCCCACTAACTCTCTCTCTGCTCTGACTCAG (SEQ ID NO: 240);

p) GTAATGATTGATTGCAATGTATGATTACAATAATCTCAGTATAAGTTCAGTAATA ATAACCTTCCACTGCTGTCCTCTGTGTGCACCCAG (SEQ ID NO: 241); or

q) GTAAATATATACAACAGTTTTTCATTTAAATAAGTGCACGGCACAAATAAGAAAA ATATGTCAAAAATGTAACCAATAGTTTTTTTCAAATTTAG (SEQ ID NO: 242).

32. The polynucleotide of any one of claims 1-31, wherein the intron comprises between about 10 base pairs to about 500 base pairs.

33. The polynucleotide of claim 32, wherein the intron comprises between about 70 base pairs and 150 base pairs.

34. The polynucleotide of claim 32, wherein the intron comprises between about 100 base pairs and 200 base pairs.

35. The polynucleotide of any one of claims 1-34, wherein the intron is inserted in proximity to a protospacer sequence.

36. The polynucleotide of claim 35, wherein the intron is inserted within about 10 to 30 base pairs of the protospacer sequence.

37. The polynucleotide of claim 35 or 36, wherein the protospacer sequence is NGG or NNGRRT.

38. The polynucleotide of any one of claims 1, 2 or 6-37, wherein the deaminase domain comprises a TadA domain.

39. The polynucleotide of claim 38, wherein the intron is inserted within or directly after codon 18, 23, 59, 62, 87, or 129 of TadA.

40. The polynucleotide of claim 39, wherein the intron is inserted directly after codon 87 of TadA.

41. The polynucleotide of any one of claims 2, 5 or 8, wherein the alteration is a single-base edit.

42. The polynucleotide of claim 41, wherein the single-base edit is an A-to-G base edit.

43. The polynucleotide of claim 41, wherein the single-base edit is a C-to-T base edit.

44. The polynucleotide of any one of claims 1-43, further comprising a polynucleotide sequence encoding a linker.

45. The polynucleotide of claim 44, wherein the intron is inserted within the polynucleotide sequence encoding the linker.

46. The polynucleotide of any one of claims 1-45, wherein the programmable DNA binding protein domain is a Cas9 domain.

47. The polynucleotide of claim 46, wherein the Cas9 domain is split between amino acid residues corresponding to Asn309 and Thr310 of Cas9, and residue 310 was mutated to a Thr310Cys.

48. A composition comprising:

(i) a first polynucleotide encoding a deaminase domain and an N- terminal fragment of a nucleic acid programmable DNA binding protein (napDNAbp) domain, wherein the N-terminal fragment of the napDNAbp domain is fused to a split intein-N, and

(ii) a second polynucleotide encoding a C-terminal fragment of the napDNAbp domain, wherein the C- terminal fragment of the napDNAbp domain is fused to a split intein-C;

wherein the first or second polynucleotide comprises an intron, wherein the intron is inserted in an open reading frame of the polynucleotides.

49. A composition comprising: (i) a first polynucleotide encoding an N-terminal fragment of a deaminase domain, wherein the N-terminal fragment of the deaminase domain is fused to a split intein-N, and (ii) a second polynucleotide encoding a C-terminal fragment of the deaminase domain and a nucleic acid programmable DNA binding protein (napDNAbp) domain, wherein the C- terminal fragment of the deaminase domain is fused to a split intein-C;

wherein the first or second polynucleotide comprises an intron, wherein the intron is inserted in an open reading frame of the polynucleotides.

50. The composition of any one of claims 48 or 49, wherein the intron comprises an alteration at a splice acceptor or splice donor site, wherein the alteration reduces or eliminates splicing of base editor mRNA.

51. The composition of any one of claims 48-50, wherein the deaminase domain is a cytidine deaminase domain or an adenosine deaminase domain.

52. The composition of any one of claims 48-51, wherein the deaminase domain is a TadA domain.

53. The composition of any one of claims 48-52, wherein the napDNAbp domain is a Cas domain selected from the group consisting of a Cas9, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, and Cas12j/CasΦ domain.

54. The composition of any one of claims 48-53, wherein the napDNAbp domain is a Cas9 domain.

55. The composition of claim 54, wherein the N- and C-terminal domains of the Cas9 domain are split between amino acid residues Asn309 and Thr310.

56. The composition of claim 54 or 55, wherein the Cas9 domain comprises the mutation Thr310Cys.

57. The composition of any one of claims 48-56, wherein the intron is derived from a sequence selected from the group consisting of NF1, PAX2, EEF1A1, HBB, IGHG1, SLC50A1, ABCB11, BRSK2, PLXNB3, TMPRSS6, IL32, ANTXRL, PKHD1L1, PADI1, KRT6C, and HMCN2.

58. The composition of claim 57, wherein the intron is derived from NF1.

59. The composition of claim 57, wherein the intron is derived from PAX2.

60. The composition of claim 57, wherein the intron is derived from EEF1A1.

61. The composition of claim 57, wherein the intron is derived from HBB.

62. The composition of claim 57, wherein the intron is derived from IGHG1.

63. The composition of claim 57, wherein the intron is derived from SLC50A1.

64. The composition of claim 57, wherein the intron is derived from ABCB11.

65. The composition of claim 57, wherein the intron is derived from BRSK2.

66. The composition of claim 57, wherein the intron is derived from PLXNB3.

67. The composition of claim 57, wherein the intron is derived from TMPRSS6.

68. The composition of claim 57, wherein the intron is derived from IL32.

69. The composition of claim 57, wherein the intron is derived from PKHD1L1.

70. The composition of claim 57, wherein the intron is derived from PADI1.

71. The composition of claim 57, wherein the intron is derived from KRT6C.

72. The composition of claim 57, wherein the intron is derived from HMCN2.

73. The composition of any one of claims 48-72, further comprising a linker polynucleotide sequence.

74. The composition of claim 73, wherein the intron is inserted within the linker polynucleotide sequence.

75. A base editor system comprising:

(i) a polynucleotide encoding a base editor comprising a deaminase domain, or fragment thereof;

(ii) one or more guide RNAs that direct the base editor to edit a site in the genome of a cell; and

(iii) one or more guide RNAs that direct the base editor to edit the polynucleotide encoding the base editor, wherein the edit results in a decrease in activity and/or expression of the encoded base editor.

76. The base editor system of claim 75, wherein the edit alters a catalytic residue of the deaminase domain.

77. The base editor of claim 75 or claim 76, wherein the deaminase domain is an adenosine deaminase domain.

78. The base editor of claim 75 or claim 76, wherein the deaminase domain is and cytidine deaminase domain.

79. The base editor system of claim 77, wherein the altered catalytic residue of the deaminase domain is His57 (H57), Glu59 (E59), Cys87 (C87), or Cys90 (C90) of the following reference sequence:
MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 1), or a corresponding position in another adenosine deaminase.

80. The base editor system of claim 76 or claim 79, wherein the altered catalytic residue is E59.

81. The base editor of claim 76 or claim 79, wherein the alteration to the catalytic residue is E59G.

82. The base editor system of claim 76 or claim 79, wherein the altered catalytic residue is H57.

83. The base editor of claim 76 or claim 79, wherein the alteration to the catalytic residue is H57R.

84. The base editor system of claim 76 or claim 79, wherein the altered catalytic residue is C87.

85. The base editor of claim 76 or claim 79, wherein the alteration to the catalytic residue is C87R.

86. The base editor system of claim 76 or claim 79, wherein the altered catalytic residue is C90.

87. The base editor of claim 76 or claim 79, wherein the alteration to the catalytic residue is C90R.

88. A base editor system comprising:

(i) a polynucleotide encoding a self-inactivating base editor or fragment thereof, wherein the polynucleotide comprises an intron inserted in an open reading frame of the self-inactivating base editor or fragment thereof;

(ii) one or more guide RNAs that direct the self-inactivating base editor to edit a site in the genome of a cell; and

(iii) one or more guide RNAs that direct the self-inactivating base editor to edit a splice acceptor or a splice donor site present in the intron of the polynucleotide encoding the self-inactivating base editor.

89. A base editor system comprising:

(i) the polynucleotide of any one of claims 3-47 encoding a base editor;

(ii) one or more guide RNAs that direct the base editor to edit a site in the genome of a cell; and

(iii) one or more guide RNAs that direct the base editor to edit a splice acceptor or a splice donor site present in the intron of the polynucleotide encoding the base editor.

90. A base editor system comprising:

(i) the composition of any one of claims 48-74 encoding a base editor;

(ii) one or more guide RNAs that direct the base editor to edit a site in the genome of a cell; and (iii) one or more guide RNAs that direct the base editor to edit a splice acceptor or a splice donor site present in the intron of the composition of (i).

91. A base editor system comprising:

(i) a first polynucleotide encoding a deaminase domain and an N-terminal fragment of a nucleic acid programmable DNA binding protein (napDNAbp) domain, wherein the N-terminal fragment of the napDNAbp domain is fused to a split intein-N;

(ii) a second polynucleotide encoding a C-terminal fragment of the napDNAbp domain, wherein the C-terminal fragment of the napDNAbp domain is fused to a split intein-C,

wherein the first or second polynucleotide comprises an intron, wherein the intron is inserted in an open reading frame, and wherein the first and second polynucleotides encode a base editor;

(iii) one or more guide RNAs that direct the base editor to edit a site in the genome of a cell; and

(iv) one or more guide RNAs that direct the base editor to edit a splice acceptor or a splice donor site present in the intron of the polynucleotide of (i) or (ii).

92. A base editor system comprising:

(i) a first polynucleotide encoding an N- terminal fragment of a deaminase domain, wherein the N-terminal fragment of the deaminase domain is fused to a split intein-N;

(ii) a second polynucleotide encoding a C-terminal fragment of the deaminase domain and a nucleic acid programmable DNA binding protein (napDNAbp) domain, wherein the C-terminal fragment of the deaminase domain is fused to a split intein-C,

wherein the first or second polynucleotide comprises an intron, wherein the intron is inserted in an open reading frame, and wherein the first and second polynucleotides encode a base editor;

(iii) one or more guide RNAs that direct the base editor to edit a site in the genome of a cell; and

(iv) one or more guide RNAs that direct the base editor to edit a splice acceptor or a splice donor site present in the intron of the polynucleotide of (i) or (ii).

93. The base editor system of any one of claims 75-92, wherein the base editor system comprises a polynucleotide sequence selected from the following:

a) gGUUUUAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 191);

b) gUUUCUUACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 192);

c) gGUUUCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 193);

d) GCCACUUACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 194);

e) gACAUUAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 195);

f) gGAUCUCACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 196);

g) gUCCUUAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 197);

h) GUCACCUACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 198);

i) GAUUUCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 190);

j) gGUGCUUACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 200);

k) gUCCACAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 201);

l) GAUACUUACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 202);

m) gUGUUUUAGCUGCGGCAAGGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 203);

n) gUUUCUUACAGCCAUAAUUUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 204);

o) gCUCCACAGCUGCGGCAAGGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 205);

p) GAUACUUACAGCCAUAAUUUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 206);

q) gUGUUUUAGGGACGAAAGAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 207);

r) gUUACCUGGCUCUCUUAGCCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 208);

s) gCUCCACAGGGACGAAAGAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 209);

t) gCUUGCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 210);

u) gAUUGCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 211);

v) gUCUCCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 212);

w) gUCUGCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 213);

x) gGACUCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 214);

y) GCACCCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 215);

z) gAAUUUAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 216);

aa) gCAUUAGGUCGAGAUCACAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 217);

bb) gCCUUAGGUCGAGAUCACAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 218);

cc) GUUUCAGGUCGAGAUCACAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 219);

dd) gACAUUAGGCUAAGAGAGCCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 220);

ee) gUCCUUAGGCUAAGAGAGCCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 221);

ff) gGUUUCAGGCUAAGAGAGCCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 222);

gg) gACAUUAGAUUAUGGCUCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 223);

hh) gUCCUUAGAUUAUGGCUCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 224);

ii) gGUUUCAGAUUAUGGCUCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 225);

jj) gCACCAUGAGCGAGGUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 524);

kk) gGCCACCAUGAGCGAGGUCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 525);

ll) GUGUCGAAGUUCGCCCUGGAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 526);

mm) gAUGCCGAGAUAAUGGCCCUCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAG GCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 527);

nn) gAUGCCGAGAUAAUGGCCCUUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 528);

oo) gAUGCCGAGAUCAUGGCACUAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 529);

pp) gAUGCCGAGAUCAUGGCACUCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 530);

qq) gAUGCCGAGAUCAUGGCACUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 531);

rr) gAUGCCGAGAUCAUGGCGCUAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 532);

ss) gAUGCCGAGAUCAUGGCGCUCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 533);

tt) gAUGCCGAGAUCAUGGCGUUAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 534);

uu) gAUGCCGAGAUUAUGGCACUAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 535);

vv) gAUGCCGAGAUUAUGGCACUCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 536);

ww) gAUGCCGAGAUUAUGGCACUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAG GCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 537);

xx) gAUGCCGAGAUUAUGGCACUUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 538);

yy) gAUGCCGAGAUUAUGGCGCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 539);

zz) gAUGCCGAGAUUAUGGCUCUAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 540);

aaa) gAUGCGGAGAUCAUGGCGCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAG GCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 541);

bbb) gAUGCUGAGAUAAUGGCCCUCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAG GCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 542);

ccc) gAACCGCACAUGCCGAAAUUAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAG GCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 543);

ddd) gGCAGGUGUCGACAUAUCUAUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAG GCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 544);

eee) gAUGCCGAAAUUAUGGCUCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAG GCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 545);

fff) gACACAUGACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 546);

or

ggg) gGCCCCAGCACACAUGACACAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAG GCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 547).

94. A vector comprising a polynucleotide encoding a self-inactivating base editor or fragment thereof, wherein the polynucleotide comprises an intron inserted in an open reading frame of the self-inactivating base editor or fragment thereof.

95. A vector comprising the polynucleotide of any one of claims 1-47 or the base editor system of any one of claims 75-93.

96. A vector comprising the first polynucleotide and/or the second polynucleotide of the composition of any one of claims 48-74.

97. The vector of any one of claims 94-96, wherein the expression vector is a mammalian expression vector.

98. The vector of any one of claims 94-97, wherein the vector is a lipid nanoparticle.

99. The vector of any one of claims 94-98, wherein the vector is a viral vector selected from the group consisting of an adeno-associated virus (AAV), retroviral vector, adenoviral vector, lentiviral vector, Sendai virus vector, and herpes virus vector.

100. The vector of claim 99, wherein the vector is an AAV vector.

101. The vector of claim 100, wherein the AAV vector is AAV2 or AAV8.

102. The vector of any one of claims 94-101, wherein the vector comprises a promoter.

103. The vector of claim 102, wherein the promoter is a CMV promoter.

104. A cell comprising a vector comprising a polynucleotide encoding a self-inactivating base editor or fragment thereof, wherein the polynucleotide comprises an intron inserted in an open reading frame of the self-inactivating base editor or fragment thereof.

105. A cell comprising the polynucleotide of any one of claims 1-47, the composition of any one of claims 48-74, the base editor system of any one of claims 75-93, or the vector of any one of claims 94-103.

106. The cell of claim 104 or 105, wherein the cell is *in vitro* or *in vivo*.

107. A pharmaceutical composition comprising the polynucleotide of any one of claims 1-47, the base editor system of any one of claims 75-93, the vector of any one of claims 94-103, or the cell of any one of claims 104-106.

108. The pharmaceutical composition of claim 107, further comprising a pharmaceutically acceptable excipient, diluent, or carrier.

109. A kit comprising the polynucleotide of any one of claims 1-47, the composition of any one of claims 48-74, the base editor system of any one of claims 75-93, the vector of any one of claims 94-103, the cell of any one of claims 104-106, or the pharmaceutical composition of claim 107 or claim 108.

110. The kit of claim 109, further comprising instructions for use thereof.

111. A method for reducing or eliminating expression of a self-inactivating base editor, the method comprising:

(a) providing a polynucleotide encoding a self-inactivating base editor or fragment thereof, wherein the polynucleotide comprises an intron inserted in an open reading frame of the self-inactivating base editor or fragment thereof; and

(b) contacting the polynucleotide with a guide RNA and a self-inactivating base editor polypeptide, wherein the guide RNA directs the base editor to edit a splice acceptor or a splice donor site of the intron, thereby generating an alteration that reduces or eliminates expression of the self-inactivating base editor.

112. A method of self-inactivating base editing, the method comprising:

(a) expressing in a cell a polynucleotide encoding a base editor comprising a deaminase domain, or fragment thereof;

(b) contacting the cell with a first guide RNA that directs the base editor to edit a site in the genome of the cell, thereby generating an alteration in the genome of the cell; and

(c) contacting the cell with a second guide RNA that directs the base editor to edit the polynucleotide encoding the base editor, wherein the edit results in a decrease in activity and/or expression of the encoded base editor, thereby generating an alteration that reduces or eliminates expression of the base editor.

113. The method of claim 112, wherein the edit alters a catalytic residue of the deaminase domain.

114. The method of claim 112 or claim 113, wherein the deaminase domain is an adenosine deaminase domain.

115. The method of claim 112 or claim 113, wherein the deaminase domain is and cytidine deaminase domain.

116. The method of claim 114, wherein the altered catalytic residue of the deaminase domain is His57 (H57), Glu59 (E59), Cys87 (C87), or Cys90 (C90) of the following reference sequence:
MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 1), or a corresponding position in another adenosine deaminase.

117. The method of claim 116, wherein the altered catalytic residue is E59.

118. The method of claim 116, wherein the alteration to the catalytic residue is E59G.

119. The method of claim 116, wherein the altered catalytic residue is H57.

120. The method of claim 116, wherein the alteration to the catalytic residue is H57R.

121. The method of claim 116, wherein the altered catalytic residue is C87.

122. The method of claim 116, wherein the alteration to the catalytic residue is C87R.

123. The method of claim 116, wherein the altered catalytic residue is C90.

124. The method of claim 116, wherein the alteration to the catalytic residue is C90R.

125. A method of self-inactivating base editing, the method comprising:

(a) expressing in a cell a polynucleotide encoding a self-inactivating base editor or fragment thereof, wherein the polynucleotide comprises an intron inserted in an open reading frame of the self-inactivating base editor or fragment thereof;

(b) contacting the cell with a first guide RNA that directs the self-inactivating base editor to edit a site in the genome of the cell, thereby generating an alteration in the genome of the cell; and

(c) contacting the cell with a second guide RNA that directs the self-inactivating base editor to edit a splice acceptor or a splice donor site present in the intron of the polynucleotide of (a), thereby generating an alteration that reduces or eliminates expression of the self-inactivating base editor.

126. A method of editing the genome of an organism, the method comprising:

(a) expressing in a cell of the organism a polynucleotide encoding a self-inactivating base editor or fragment thereof, wherein the polynucleotide comprises an intron inserted in an open reading frame of the self-inactivating base editor or fragment thereof;

(b) contacting the cell with a first guide RNA that directs the self-inactivating base editor to edit a site in the genome of the cell, thereby generating an alteration in the genome of the cell; and

(c) contacting the cell with a second guide RNA that directs the self-inactivating base editor to edit a splice acceptor or a splice donor site present in the intron of the polynucleotide of (a), thereby generating an alteration that reduces or eliminates expression of the self-inactivating base editor.

127. A method of treating a subject, the method comprising:

(a) expressing in a cell of the subject a polynucleotide encoding a self-inactivating base editor or fragment thereof, wherein the polynucleotide comprises an intron inserted in an open reading frame of the self-inactivating base editor or fragment thereof;

(b) contacting the cell with a first guide RNA that directs the self-inactivating base editor to edit a site in the genome of the cell, thereby generating an alteration in the genome of the cell to treat the subject; and

(c) contacting the cell with a second guide RNA that directs the self-inactivating base editor to edit a splice acceptor or a splice donor site present in the intron of the polynucleotide of (a), thereby generating an alteration that reduces or eliminates expression of the self-inactivating base editor.

128. A method of treating a subject, the method comprising administering to the subject the base editor system of any one of claims —75-93, the vector of any one of claims 94-103,

the cell of any one of claims 104-106, or the pharmaceutical composition of claim 107 or claim 108, thereby treating the subject.

129. A method of editing the genome of an organism, the method comprising:

(a) expressing in a cell of the organism a first polynucleotide encoding a deaminase domain and an N-terminal fragment of a nucleic acid programmable DNA binding protein (napDNAbp) domain, wherein the N-terminal fragment of the napDNAbp domain is fused to a split intein-N, and a second polynucleotide encoding a C-terminal fragment of the napDNAbp domain, wherein the C- terminal fragment of the napDNAbp domain is fused to a split intein-C, wherein the first or second polynucleotide comprises an intron, wherein the intron is inserted in an open reading frame, and wherein expression of the first and second polynucleotides in the cell result in the formation of a self-inactivating base editor;

(b) contacting the cell with a first guide RNA that directs the self-inactivating base editor to edit a site in the genome of the cell, thereby generating an alteration in the genome of the cell; and

(c) contacting the cell with a second guide RNA that directs the self-inactivating base editor to edit a splice acceptor or a splice donor site present in the intron of the polynucleotide of (a), thereby generating an alteration that reduces or eliminates expression of the self-inactivating base editor.

130. A method of editing the genome of an organism, the method comprising:

(a) expressing in a cell of the organism a first polynucleotide encoding an N-terminal fragment of a deaminase domain, wherein the N-terminal fragment of the deaminase domain is fused to a split intein-N, and a second polynucleotide encoding a C-terminal fragment of the deaminase domain and a nucleic acid programmable DNA binding protein (napDNAbp) domain, wherein the C- terminal fragment of the deaminase domain is fused to a split intein-C, wherein the first or second polynucleotide comprises an intron, wherein the intron is inserted in an open reading frame, and wherein expression of the first and second polynucleotides in the cell result in the formation of a self-inactivating base editor;

(b) contacting the cell with a first guide RNA that directs the self-inactivating base editor to edit a site in the genome of the cell, thereby generating an alteration in the genome of the cell; and

(c) contacting the cell with a second guide RNA that directs the self-inactivating base editor to edit a splice acceptor or a splice donor site present in the intron of the polynucleotide of

(a), thereby generating an alteration that reduces or eliminates expression of the self-inactivating base editor.

131. The method of any one of claims 111-130, wherein the method is performed *in vivo*.

132. The method of any one of claims 129-130, wherein the first polynucleotide and/or second polynucleotide are expressed in a cell by a vector.

133. The method of any one of claims 129-130, wherein the first polynucleotide and second polynucleotide are expressed in a cell by separate vectors.

134. The method of any one of claims 112-133, wherein the first guide RNA and/or second guide RNA are delivered to the cell by a vector.

135. The method of any one of claims 112-133, wherein the first guide RNA and/or second guide RNA are delivered to the cell in the same vector than the first polynucleotide and/or second polynucleotide.

136. The method of any one of claims 129-135, wherein the first guide RNA and/or second guide RNA are delivered to the cell in a different vector than the first polynucleotide and/or second polynucleotide.

137. The method of any one of claims 132-136, wherein the vector is a lipid nanoparticle.

138. The method of any one of claims 132-137, wherein the vector is a viral vector.

139. The method of claim 138, wherein the viral vector is an adeno-associated virus (AAV) vector.

140. The method of claim 139, wherein the AAV vector is AAV2 or AAV8.

141. The method of any one of claims 129-140, wherein the napDNAbp domain is a Cas9 domain.

142. The method of claim 141, wherein the N- and C-terminal domains of the Cas9 domain are split between amino acid residues Asn309 and Thr310.

143. The method of claim 141 or 142, wherein the Cas9 domain comprises the mutation Thr310Cys.

144. The method of any one of claims 111-143, wherein the base editor comprises a nucleic acid programmable DNA binding protein (napDNAbp) domain and a deaminase domain.

145. The method of claim 144, wherein the open reading frame comprising the intron is in the napDNAbp domain or the deaminase domain.

146. The method of any one of claims 11 or 125-145, wherein the self-inactivating base editor polypeptide maintains high editing efficiency in genomic DNA.

147. The method of any one of claims 83, 84, 112-124, or 129-146 wherein the deaminase domain is a cytidine deaminase domain or an adenosine deaminase domain.

148. The method of claim 144 or claim 145, wherein the napDNAbp domain is a Cas domain selected from the group consisting of a Cas9, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, and Cas12j/CasΦ domain.

149. The method of any one of claims 111, 125-127, or 129-148, wherein the alteration is in a consensus splice donor site at the 5' end of the intron or in a consensus splice acceptor sequence at the 3' end of the intron.

150. The method of any one of claims 111, 125-127, or 129-149, wherein the intron comprises between about 10 base pairs to about 500 base pairs.

151. The method of claim 150, wherein the intron comprises between about 70 base pairs and 150 base pairs.

152. The method of claim 150, wherein the intron comprises between about 100 base pairs and 200 base pairs.

153. The method of any one of claims 111, 125-127 or 129-152, wherein the intron is inserted in proximity to a protospacer sequence.

154. The method of claim 153, wherein the intron is inserted within about 10 to 30 base pairs of the protospacer sequence.

155. The method of claim 153 or 154, wherein the protospacer sequence is NGG or NNGRRT.

156. The method of claim 147, wherein the adenosine deaminase domain comprises a TadA domain.

157. The method of claim 156, wherein the intron is inserted within or directly after codon 18, 23, 59, 62, 87, or 129 of TadA.

158. The method of claim 157, wherein the intron is inserted directly after codon 87 of TadA.

159. The method of any one of claims 111-127 or 129-158, wherein the alteration is a single-base edit.

160. The method of claim 159, wherein the single-base edit is an A-to-G base edit.

161. The method of claim 159, wherein the single-base edit is a C-to-T base edit.

162. The method of any one of claims 111, 125-127, or 129-161, wherein the intron is derived from a sequence selected from the group consisting of NF1, PAX2, EEF1A1, HBB, IGHG1, SLC50A1, ABCB11, BRSK2, PLXNB3, TMPRSS6, IL32, ANTXRL, PKHD1L1, PADI1, KRT6C, and HMCN2.

163. The method of claim 162, wherein the intron is derived from NF1.

164. The method of claim 162, wherein the intron is derived from PAX2.

165. The method of claim 162, wherein the intron is derived from EEF1A1.

166. The method of claim 162, wherein the intron is derived from HBB.

167. The method of claim 162, wherein the intron is derived from IGHG1.

168. The method of claim 162, wherein the intron is derived from SLC50A1.

169. The method of claim 162, wherein the intron is derived from ABCB11.

170. The method of claim 162, wherein the intron is derived from BRSK2.

171. The method of claim 162, wherein the intron is derived from PLXNB3.

172. The method of claim 162, wherein the intron is derived from TMPRSS6.

173. The method of claim 162, wherein the intron is derived from IL32.

174. The method of claim 162, wherein the intron is derived from PKHD1L1.

175. The method of claim 162, wherein the intron is derived from PADI1.

176. The method of claim 162, wherein the intron is derived from KRT6C.

177. The method of claim 162, wherein the intron is derived from HMCN2.

178. The method of any one of claims 111, 125-127, or 129-161, wherein the intron has at least about 85% nucleic acid sequence identity to an intron naturally present in a mammalian gene.

179. The method of any one of claims 111, 125-127, or 129-161, wherein the intron has at least about 85% nucleic acid sequence identity to an intron naturally present in a non-mammalian gene.

180. The method of any one of claims 111, 125-127, or 129-161, wherein the intron is a synthetic intron.

181. The method of any one of claims 111, 125-127, or 129-161, wherein the intron comprises a sequence that has at least about 85%, 90%, 95%, or 99% nucleic acid sequence identity to one of the following:

a) GTGAGATCAAATGAAAGTTTCATATAGAAATACAAAACCTAGAGAACTGGCATGT AAGAGAAGCAAAAATTACTTCAGCAAGGCCATGTTAGTAAATTTGCATCTGTTTGTCCACAT TAG (SEQ ID NO: 226);

b) GTAGGTGACAATGCTGCAGCTGCCTAATCTAGGTGGGGGGAACTAAATTGTGGGT GAGCTGCTGAATGGTCTGTAGTCTGAGGCTGGGGTGGGGGGAGACACAACGTCCCCTCCCTG CAAACCACTGCTATTCTGTCCCTCTCTCTCCTTAG (SEQ ID NO: 227);

c) GTAAGTGGCTTTCAAGACCATTGTTAAAAAGCTCTGGGAATGGCGATTTCATGCT TACATAAATTGGCATGCTTGTGTTTCAG (SEQ ID NO: 228);

d) GTAAGTATCAAGGTTACAAGACAGGTTTAAGGAGACCAATAGAAACTGGGCTTGT CTAGACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTACTGACATCCACTTT GCCTTTCTCTCCACAG (SEQ ID NO: 229);

e) GTAAGCACAACTGGGATGGGGTGACAGGGGTGCAAGATTGAAAACTGGCTCCTCT CCTCATAGCAGTTCTTGTGATTTCAG (SEQ ID NO: 230);

f) GTAAGAAATGTTATTTTTCAGTAAGTGATTTAGTTATTTTTCCTTTTTTTCTCATTA AAATTTCTCTAACATCTCCCTCTTCATGTTTTAG (SEQ ID NO: 231);

g) GTGAGACCCTAGCCCCCTCAACCCTGCCCTGGCCTCTCCCCAAACCTGCCCCCCC ACGCTGACCCCCACACCCGGCCGCCCGCAG (SEQ ID NO: 232);

h) GTGGGTGTCAGAGGCATCGGGGCTGCGGGGTAGGGGGCTGCCCCACCCCTAACGA AGTCTGCTCCTCCAG (SEQ ID NO: 233);

i) GCAGGGAAGTCCTGCTTCCGTGCCCCACCGGTGCTCAGCTGAGGCTCCCTTGAAAA TGCGAGGCTGTTTCCAACTTTGGTCTGTTTCCCTGGCAG (SEQ ID NO: 234);

j) GTGGGGAGTTGGGGTCCCCGAAGGTGAGGACCCTCTGGGGATGAGGGTGCTTCTCT GAGACACTTTCTTTTCCTCACACCTGTTCCTCGCCAGCAG (SEQ ID NO: 235);

k) GTATAGACCCCTTGATCTCCTAACCCTAACCCTAACCCTAACCCTAACCTACAAA ATCTTAGAGCATCAGTGGGAGCATCTCACTGTCCAGGCTCAATATTTCTTCATTTTCTTGCA G (SEQ ID NO: 236);

l) GTAATTATGATTAAAGATGGTGATTGTTTATTTTCTTTTATGATTGTCCTTAGTAT TATGTAACCTGCAAATTCTATTGCAG (SEQ ID NO: 237);

m) GTGAGTGACACAAGGTGTTGTCTGGGGAGTGGGGAAGGGGGATGGAAGTGAATCC TGTTGGTGGGGTGGAGAAAGGGCGATCTCAAGAGGGCCACTCTCTCCAG (SEQ ID NO: 238);

n) GTAAGCATCTCCACCATCCTTCTGTTTACTCTGATGGGGTCTGCAAAGGGGAGAT GATGTATAGGGTTGGGTATCTCTGTAAATGTCAGATGTGAAGTTGATCTTATGACCTTCTGT TCTGCAG (SEQ ID NO: 239);

o) GTGAGGGTCTCCCAGGCTGGGCAGGGGGAGGGGGCTGCTGCCTTGATTGCGTCCC AGGACACAGCCCTCCTCCAGCCTGCCCTCGCCTTGCTCATCCCCTCCCCATCTCAGCCCCAC CCCCACTAACTCTCTCTCTGCTCTGACTCAG (SEQ ID NO: 240);

p) GTAATGATTGATTGCAATGTATGATTACAATAATCTCAGTATAAGTTCAGTAATA ATAACCTTCCACTGCTGTCCTCTGTGTGCACCCAG (SEQ ID NO: 241); or

q) GTAAATATATACAACAGTTTTTCATTTAAATAAGTGCACGGCACAAATAAGAAAA ATATGTCAAAAATGTAACCAATAGTTTTTTTCAAATTTAG (SEQ ID NO: 242).

182. The method of any one of claims 111, 125-127, or 129-161, wherein the intron comprises a nucleic acid sequence from one of the following:

a) GTGAGATCAAATGAAAGTTTCATATAGAAATACAAAACCTAGAGAACTGGCATGT AAGAGAAGCAAAAATTACTTCAGCAAGGCCATGTTAGTAAATTTGCATCTGTTTGTCCACAT TAG (SEQ ID NO: 226);

b) GTAGGTGACAATGCTGCAGCTGCCTAATCTAGGTGGGGGGAACTAAATTGTGGGT GAGCTGCTGAATGGTCTGTAGTCTGAGGCTGGGGTGGGGGGAGACACAACGTCCCCTCCCTG CAAACCACTGCTATTCTGTCCCTCTCTCTCCTTAG (SEQ ID NO: 227);

c) GTAAGTGGCTTTCAAGACCATTGTTAAAAAGCTCTGGGAATGGCGATTTCATGCT TACATAAATTGGCATGCTTGTGTTTCAG (SEQ ID NO: 228);

d) GTAAGTATCAAGGTTACAAGACAGGTTTAAGGAGACCAATAGAAACTGGGCTTGT CTAGACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTACTGACATCCACTTT GCCTTTCTCTCCACAG (SEQ ID NO: 229);

e) GTAAGCACAACTGGGATGGGGTGACAGGGGTGCAAGATTGAAAACTGGCTCCTCT CCTCATAGCAGTTCTTGTGATTTCAG (SEQ ID NO: 230);

f) GTAAGAAATGTTATTTTTCAGTAAGTGATTTAGTTATTTTTCCTTTTTTTCTCATTA AAATTTCTCTAACATCTCCCTCTTCATGTTTTAG (SEQ ID NO: 231);

g) GTGAGACCCTAGCCCCCTCAACCCTGCCCTGGCCTCTCCCCAAACCTGCCCCCCC ACGCTGACCCCCACACCCGGCCGCCCGCAG (SEQ ID NO: 232);

h) GTGGGTGTCAGAGGCATCGGGGCTGCGGGGTAGGGGGCTGCCCCACCCCTAACGA AGTCTGCTCCTCCAG (SEQ ID NO: 233);

i) GCAGGGAAGTCCTGCTTCCGTGCCCCACCGGTGCTCAGCTGAGGCTCCCTTGAAAA TGCGAGGCTGTTTCCAACTTTGGTCTGTTTCCCTGGCAG (SEQ ID NO: 234);

j) GTGGGGAGTTGGGGTCCCCGAAGGTGAGGACCCTCTGGGGATGAGGGTGCTTCTCT GAGACACTTTCTTTTCCTCACACCTGTTCCTCGCCAGCAG (SEQ ID NO: 235);

k) GTATAGACCCCTTGATCTCCTAACCCTAACCCTAACCCTAACCCTAACCTACAAA ATCTTAGAGCATCAGTGGGAGCATCTCACTGTCCAGGCTCAATATTTCTTCATTTTCTTGCA G (SEQ ID NO: 236);

l) GTAATTATGATTAAAGATGGTGATTGTTTATTTTCTTTTATGATTGTCCTTAGTAT TATGTAACCTGCAAATTCTATTGCAG (SEQ ID NO: 237);

m) GTGAGTGACACAAGGTGTTGTCTGGGGAGTGGGGAAGGGGGATGGAAGTGAATCC TGTTGGTGGGGTGGAGAAAGGGCGATCTCAAGAGGGCCACTCTCTCCAG (SEQ ID NO: 238);

n) GTAAGCATCTCCACCATCCTTCTGTTTACTCTGATGGGGTCTGCAAAGGGGAGAT GATGTATAGGGTTGGGTATCTCTGTAAATGTCAGATGTGAAGTTGATCTTATGACCTTCTGT TCTGCAG (SEQ ID NO: 239);

o) GTGAGGGTCTCCCAGGCTGGGCAGGGGGAGGGGGCTGCTGCCTTGATTGCGTCCC AGGACACAGCCCTCCTCCAGCCTGCCCTCGCCTTGCTCATCCCCTCCCCATCTCAGCCCCAC CCCCACTAACTCTCTCTCTGCTCTGACTCAG (SEQ ID NO: 240);

p) GTAATGATTGATTGCAATGTATGATTACAATAATCTCAGTATAAGTTCAGTAATA ATAACCTTCCACTGCTGTCCTCTGTGTGCACCCAG (SEQ ID NO: 241); or

q) GTAAATATATACAACAGTTTTTCATTTAAATAAGTGCACGGCACAAATAAGAAAA ATATGTCAAAAATGTAACCAATAGTTTTTTTCAAATTTAG (SEQ ID NO: 242).

183. The method of any one of claims 112-182, wherein the second guide RNA comprises a polynucleotide sequence selected from the following:

a) gGUUUUAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 191);

b) gUUUCUUACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 192);

c) gGUUUCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 193);

d) GCCACUUACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 194);

e) gACAUUAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 195);

f) gGAUCUCACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 196);

g) gUCCUUAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 197);

h) GUCACCUACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 198);

i) GAUUUCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 190);

j) gGUGCUUACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 200);

k) gUCCACAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 201);

l) GAUACUUACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 202);

m) gUGUUUUAGCUGCGGCAAGGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 203);

n) gUUUCUUACAGCCAUAAUUUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 204);

o) gCUCCACAGCUGCGGCAAGGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 205);

p) GAUACUUACAGCCAUAAUUUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 206);

q) gUGUUUUAGGGACGAAAGAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 207);

r) gUUACCUGGCUCUCUUAGCCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 208);

s) gCUCCACAGGGACGAAAGAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 209);

t) gCUUGCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 210);

u) gAUUGCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 211);

v) gUCUCCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 212);

w) gUCUGCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 213);

x) gGACUCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 214);

y) GCACCCAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 215);

z) gAAUUUAGGUCAUGUGUGCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 216);

aa) gCAUUAGGUCGAGAUCACAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 217);

bb) gCCUUAGGUCGAGAUCACAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 218);

cc) GUUUCAGGUCGAGAUCACAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 219);

dd) gACAUUAGGCUAAGAGAGCCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 220);

ee) gUCCUUAGGCUAAGAGAGCCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 221);

ff) gGUUUCAGGCUAAGAGAGCCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 222);

gg) gACAUUAGAUUAUGGCUCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 223);

hh) gUCCUUAGAUUAUGGCUCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 224);

ii) gGUUUCAGAUUAUGGCUCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 225);

jj) gCACCAUGAGCGAGGUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 524);

kk) gGCCACCAUGAGCGAGGUCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 525);

ll) GUGUCGAAGUUCGCCCUGGAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 526);

mm) gAUGCCGAGAUAAUGGCCCUCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAG GCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 527);

nn) gAUGCCGAGAUAAUGGCCCUUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 528);

oo) gAUGCCGAGAUCAUGGCACUAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 529);

pp) gAUGCCGAGAUCAUGGCACUCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 530);

qq) gAUGCCGAGAUCAUGGCACUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 531);

rr) gAUGCCGAGAUCAUGGCGCUAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 532);

ss) gAUGCCGAGAUCAUGGCGCUCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 533);

tt) gAUGCCGAGAUCAUGGCGUUAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 534);

uu) gAUGCCGAGAUUAUGGCACUAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 535);

vv) gAUGCCGAGAUUAUGGCACUCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 536);

ww) gAUGCCGAGAUUAUGGCACUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAG GCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 537);

xx) gAUGCCGAGAUUAUGGCACUUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 538);

yy) gAUGCCGAGAUUAUGGCGCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 539);

zz) gAUGCCGAGAUUAUGGCUCUAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 540);

aaa) gAUGCGGAGAUCAUGGCGCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAG GCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 541);

bbb) gAUGCUGAGAUAAUGGCCCUCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAG GCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 542);

ccc) gAACCGCACAUGCCGAAAUUAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAG GCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 543);

ddd) gGCAGGUGUCGACAUAUCUAUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAG GCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 544);

eee) gAUGCCGAAAUUAUGGCUCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAG GCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 545);

fff) gACACAUGACACAGGGCUCGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 546);

or

ggg) gGCCCCAGCACACAUGACACAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAG GCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (SEQ ID NO: 547).

184. The method of any one of claims 111-127 or 129-183, wherein the polynucleotide further comprises a linker polynucleotide sequence.

185. The method of claim 184, wherein the intron is inserted within the linker polynucleotide sequence.

186. The method of claims 126-130, wherein the subject or organism is human.

187. The method of claim 186, wherein the subject or organism is a mammal.

188. The method of claim 187, wherein the mammal is human.

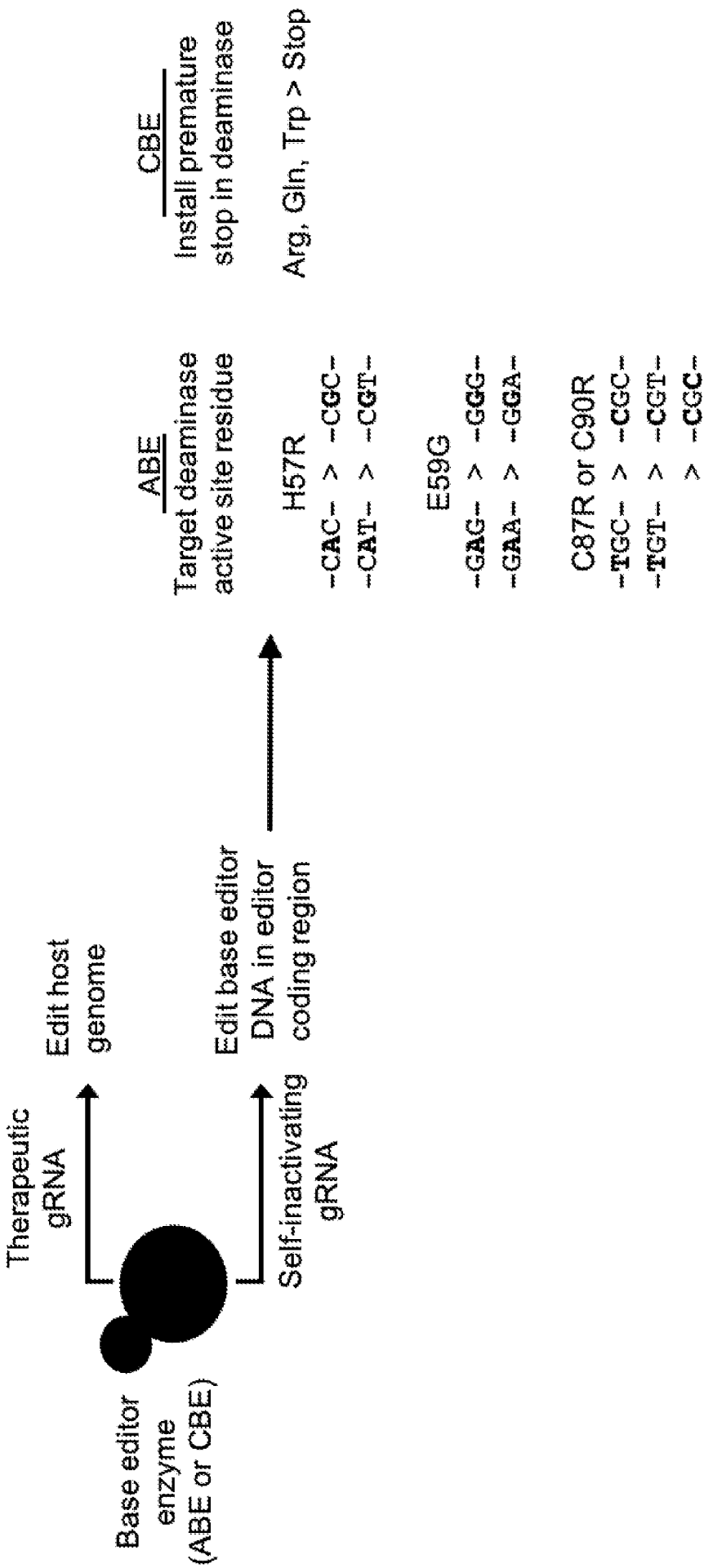

Base editor enzyme (ABE or CBE)
Therapeutic gRNA
Edit host genome
Self-inactivating gRNA
Edit base editor DNA in editor coding region
ABE
Target deaminase active site residue
H57R
-CAC- > -CGC-
-CAT- > -CGT-
E59G
-GAG- > -GGG-
-GAA- > -GGA-
C87R or C90R
-TGC- > -CGC-
-TGT- > -CGT-
> -CGC-
CBE
Install premature stop in deaminase
Arg, Gln, Trp > Stop

FIG. 1A

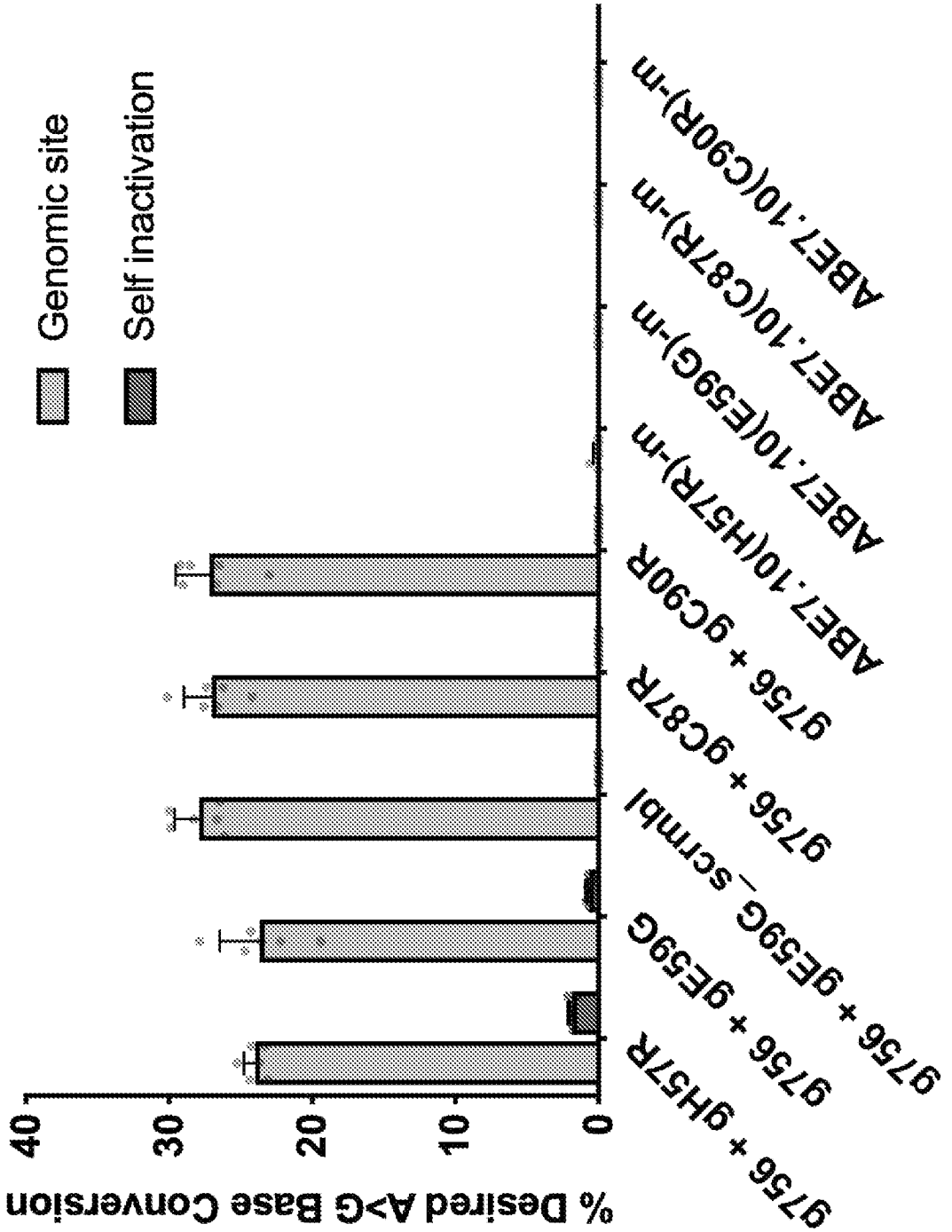

% Desired A>G Base Conversion
0
10
20
30
40
Genomic site
Self inactivation
9756 + gH57R
9756 + gE59G
9756 + gE59G_scrmbl
9756 + gC87R
9756 + gC90R
ABE7.10(H57R)-m
ABE7.10(E59G)-m
ABE7.10(C87R)-m
ABE7.10(C90R)-m

FIG. 1B

1st gen H57 target site:
1 9 20
AACCGCACATGCCGAAATTATGG
T A H A E I M
1st gen E59 target site:
1 7 20
ATGCCGAAATTATGGCTCTGGGG
A E I M A L R

FIG. 1C

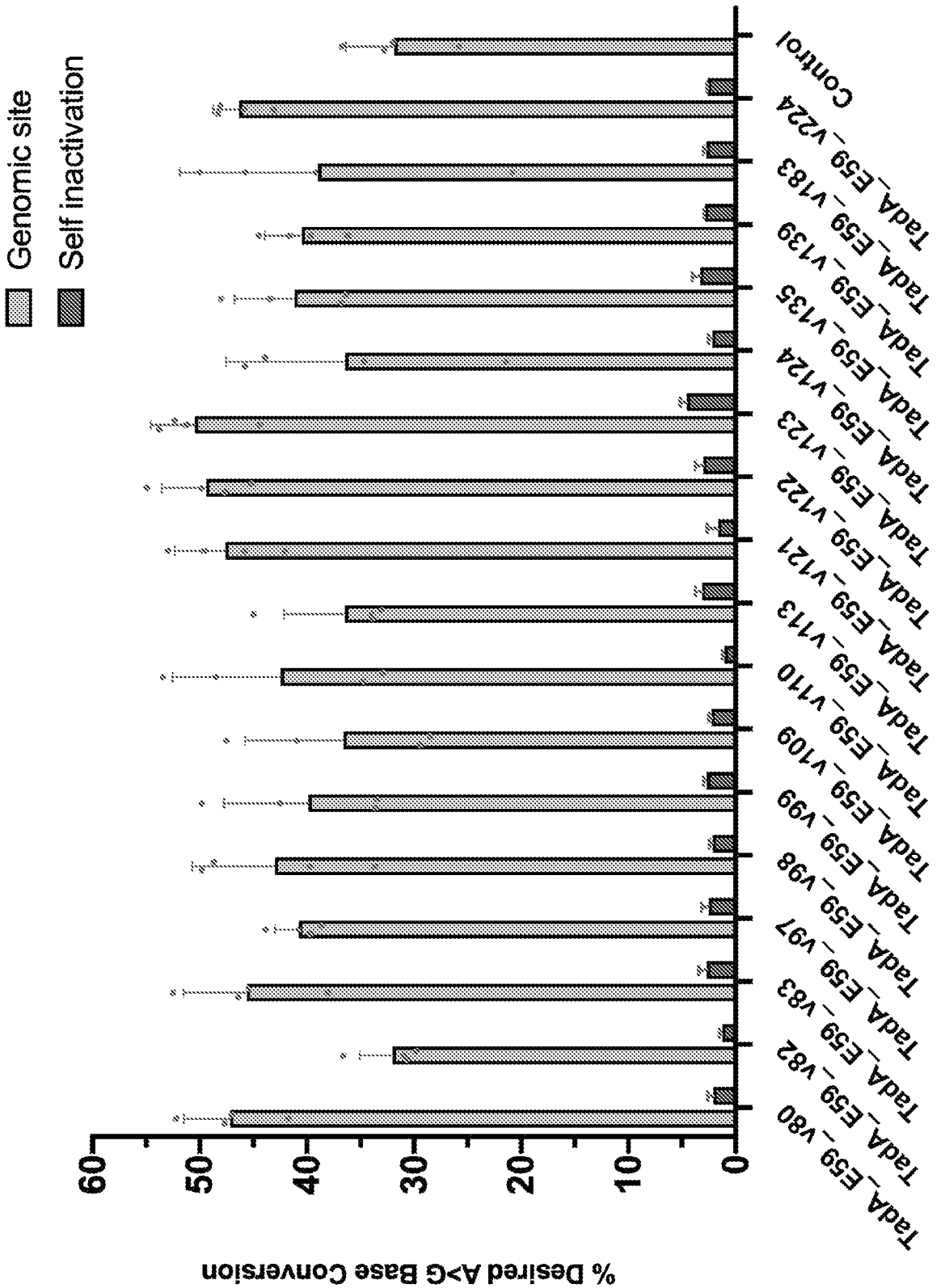

Genomic site
Self inactivation
% Desired A>G Base Conversion
0
10
20
30
40
50
60
TadA_E59_V80
TadA_E59_V82
TadA_E59_V83
TadA_E59_V97
TadA_E59_V98
TadA_E59_V99
TadA_E59_V109
TadA_E59_V110
TadA_E59_V113
TadA_E59_V121
TadA_E59_V122
TadA_E59_V123
TadA_E59_V124
TadA_E59_V135
TadA_E59_V139
TadA_E59_V183
TadA_E59_V22A
Control

FIG. 1D

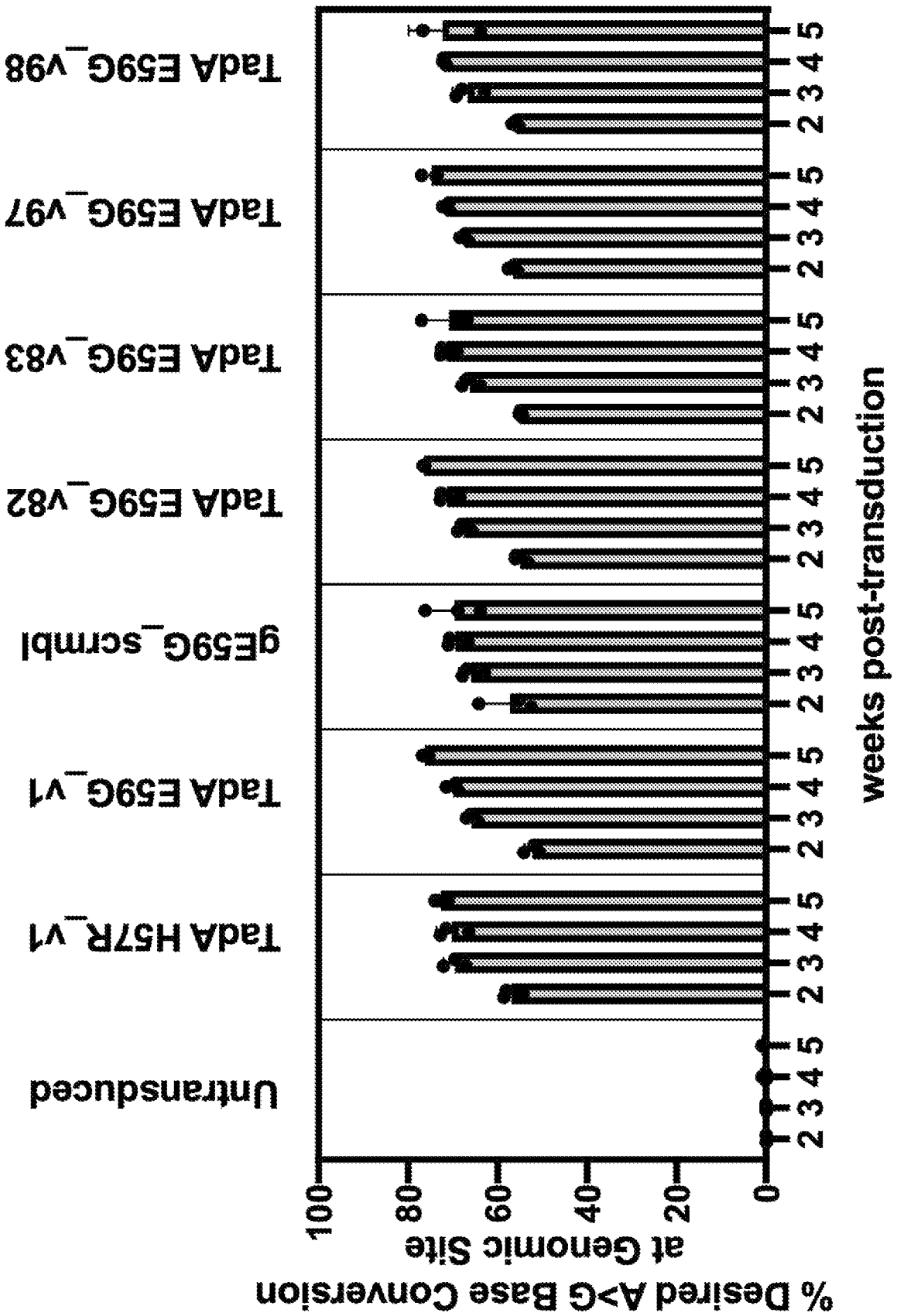

% Desired A>G Base Conversion at Genomic Site
100
80
60
40
20
0
weeks post-transduction
1 2 3 4 5
Untransduced
TadA H57R_v1
TadA E59G_v1
gE59G_scrmbl
TadA E59G_v82
TadA E59G_v83
TadA E59G_v97
TadA E59G_v98

FIG. 1E

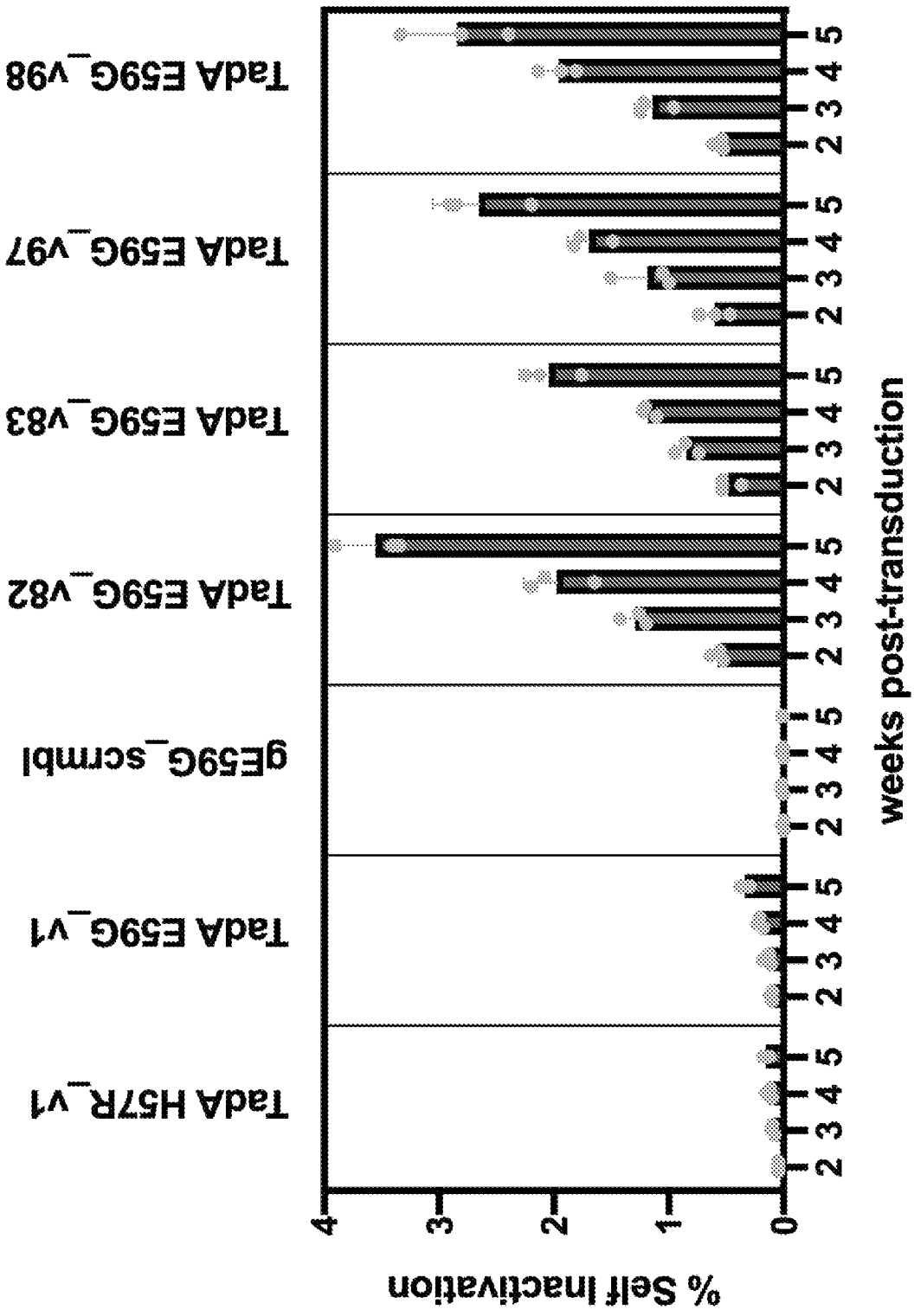

% Self Inactivation
0
1
2
3
4
weeks post-transduction
1 2 3 4 5
TadA H57R_v1
TadA E59G_v1
gE59G_scrmbl
TadA E59G_v82
TadA E59G_v83
TadA E59G_v97
TadA E59G_v98

FIG. 1F

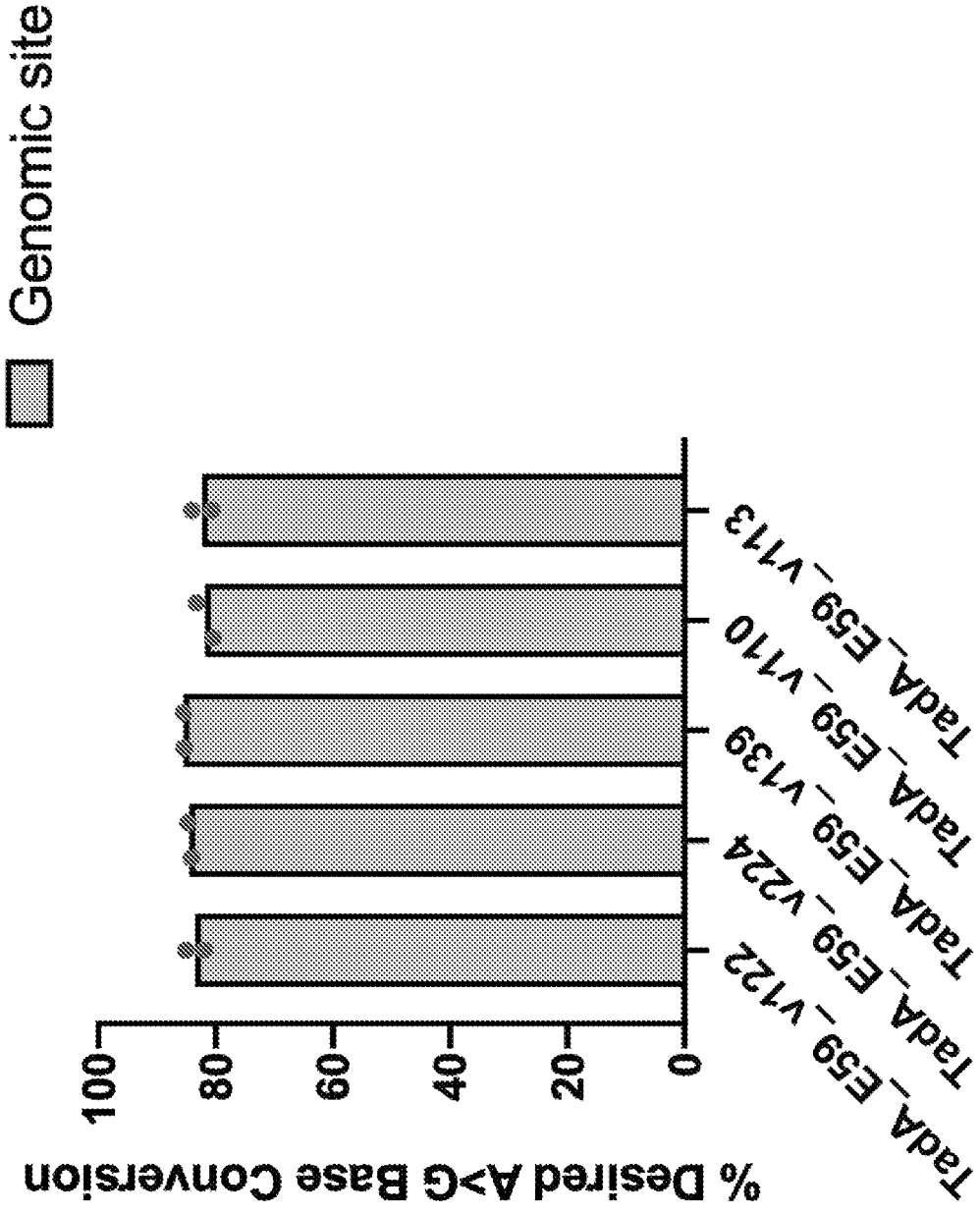

Genomic site
% Desired A>G Base Conversion
100
80
60
40
20
0
TadA_E59_V122
TadA_E59_V224
TadA_E59_V139
TadA_E59_V110
TadA_E59_V113

FIG. 1G

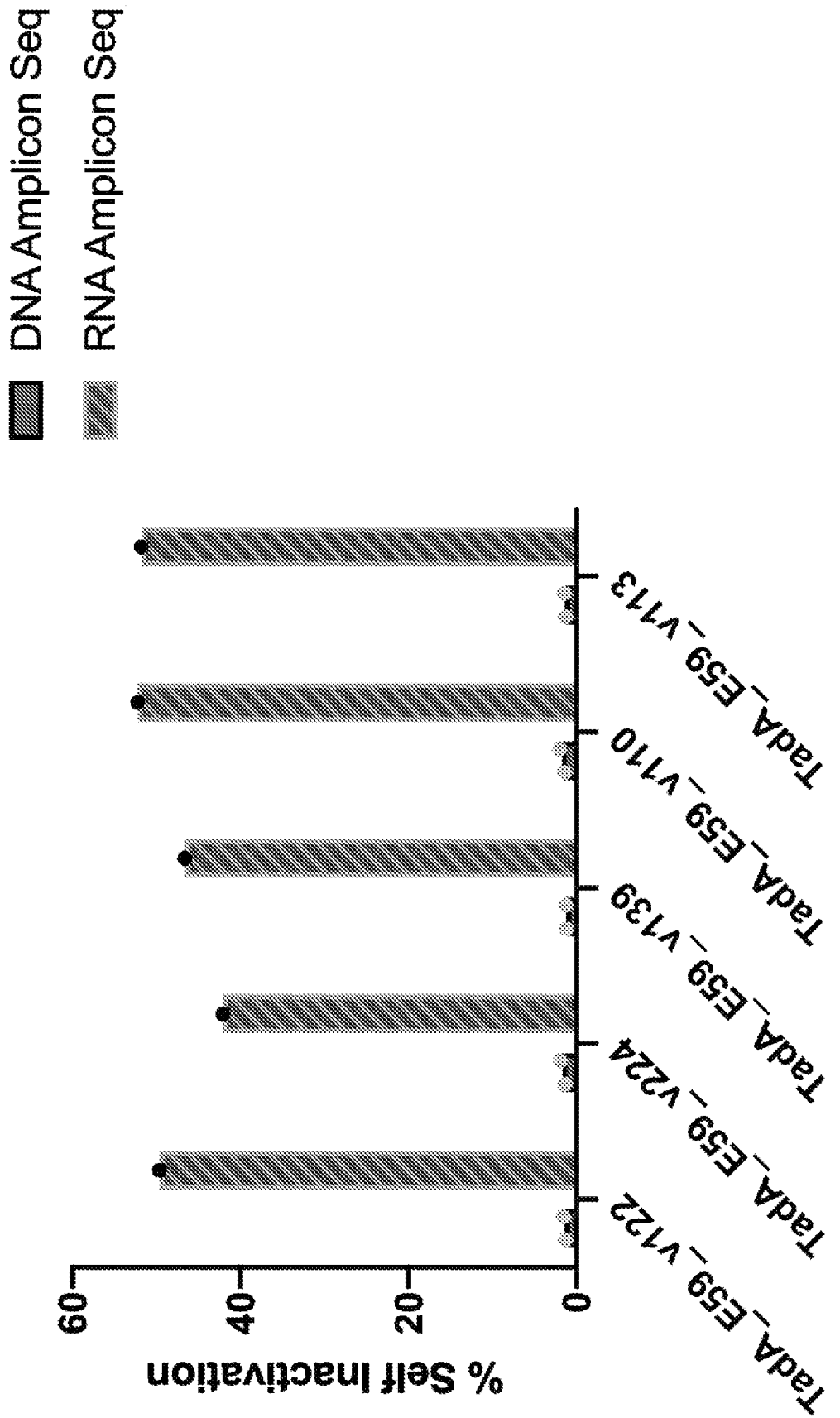

DNA Amplicon Seq
RNA Amplicon Seq
% Self Inactivation
0
20
40
60
TadA_E59_V122
TadA_E59_V224
TadA_E59_V139
TadA_E59_V110
TadA_E59_V113

FIG. 1H

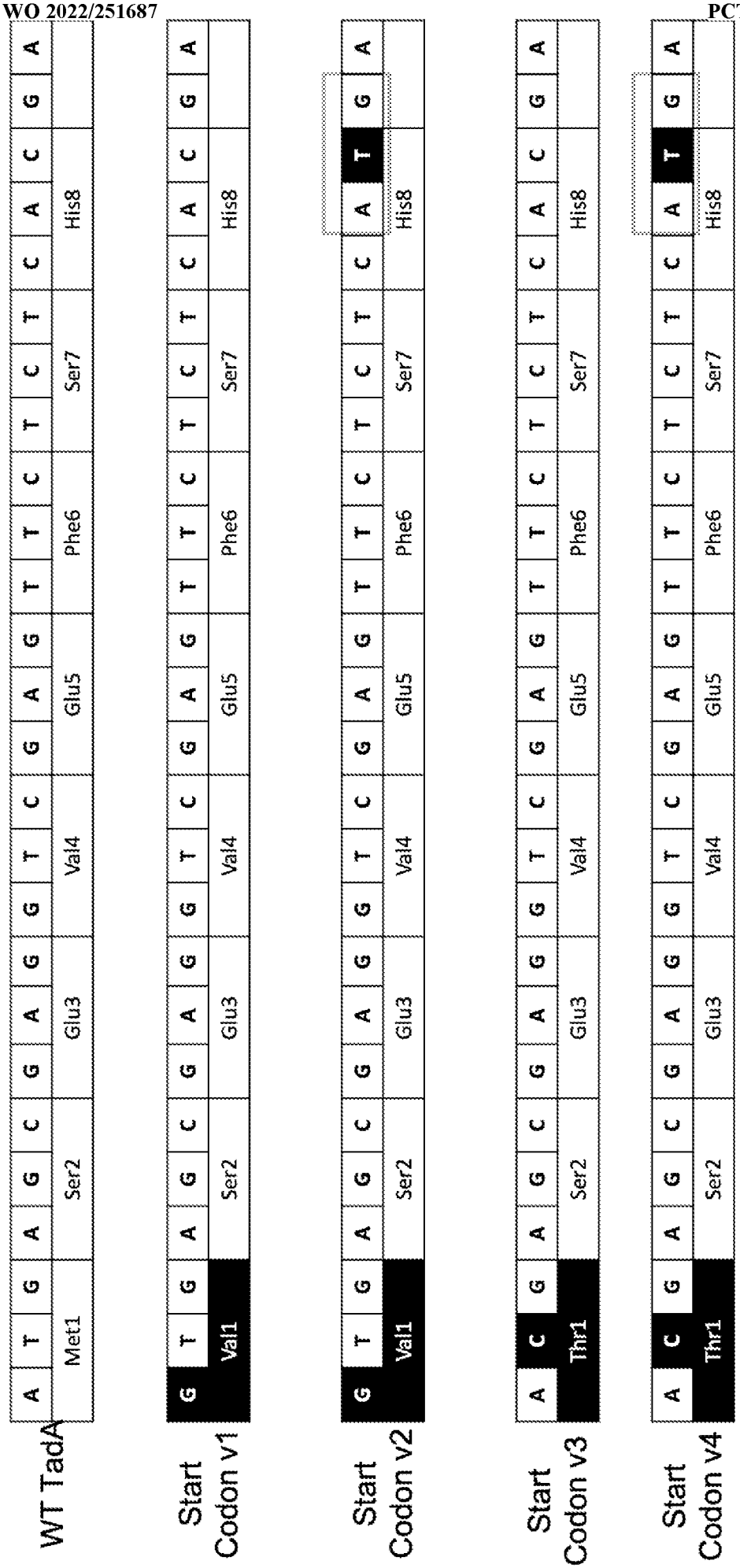

WT TadA
Met1 Ser2 Glu3 Val4 Glu5 Phe6 Ser7 His8
Start Codon v1
Val1 Ser2 Glu3 Val4 Glu5 Phe6 Ser7 His8
Start Codon v2
Val1 Ser2 Glu3 Val4 Glu5 Phe6 Ser7 His8
Start Codon v3
Thr1 Ser2 Glu3 Val4 Glu5 Phe6 Ser7 His8
Start Codon v4
Thr1 Ser2 Glu3 Val4 Glu5 Phe6 Ser7 His8

FIG. 2A

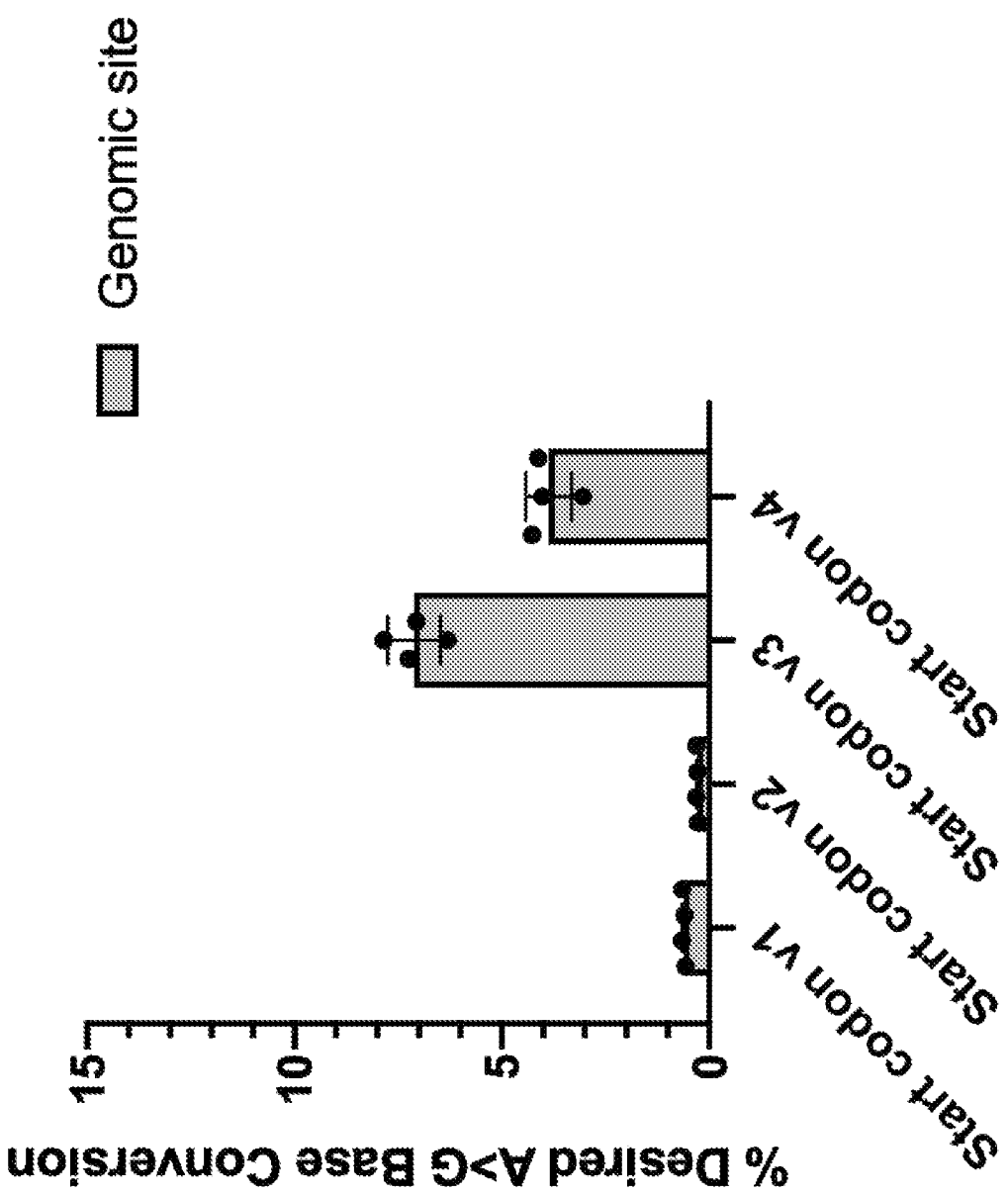

Genomic site
% Desired A>G Base Conversion
15
10
5
0
Start codon V1
Start codon V2
Start codon V3
Start codon V4

FIG. 2B

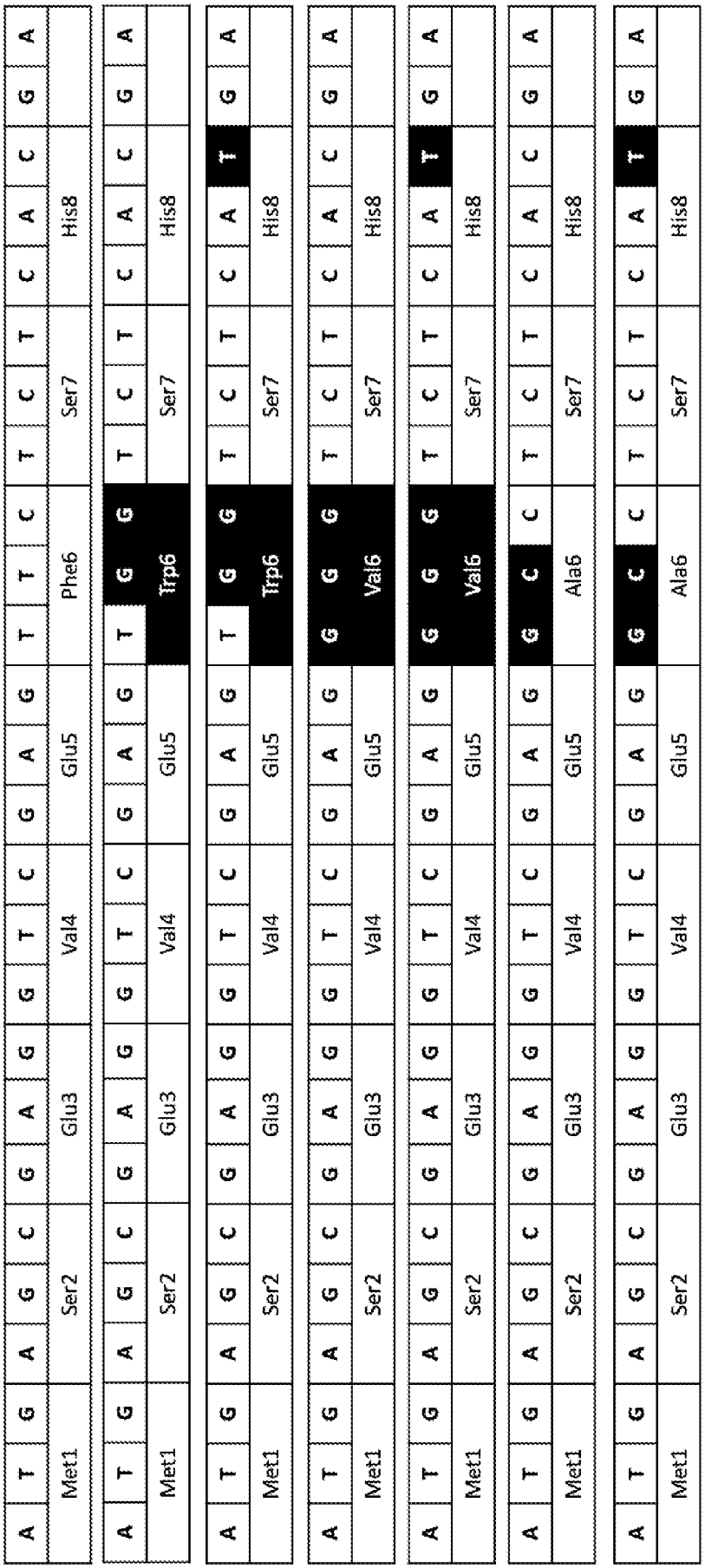

WT TadA
A T G A G C G A G G T C G A G T T C T C C A C G A
Met1 Ser2 Glu3 Val4 Glu5 Phe6 Ser7 His8
Start Codon v5
A T G A G C G A G G T C G A G T G G G T C C A C G A
Met1 Ser2 Glu3 Val4 Glu5 Trp6 Ser7 His8
Start Codon v6
A T G A G C G A G G T C G A G T G G G T C C A T G A
Met1 Ser2 Glu3 Val4 Glu5 Trp6 Ser7 His8
Start Codon v7
A T G A G C G A G G T C G A G G G G T C C A C G A
Met1 Ser2 Glu3 Val4 Glu5 Val6 Ser7 His8
Start Codon v8
A T G A G C G A G G T C G A G G G G T C C A T G A
Met1 Ser2 Glu3 Val4 Glu5 Val6 Ser7 His8
Start Codon v9
A T G A G C G A G G T C G A G G C C T C C A C G A
Met1 Ser2 Glu3 Val4 Glu5 Ala6 Ser7 His8
Start Codon v10
A T G A G C G A G G T C G A G G C C T C C A T G A
Met1 Ser2 Glu3 Val4 Glu5 Ala6 Ser7 His8

FIG. 2C

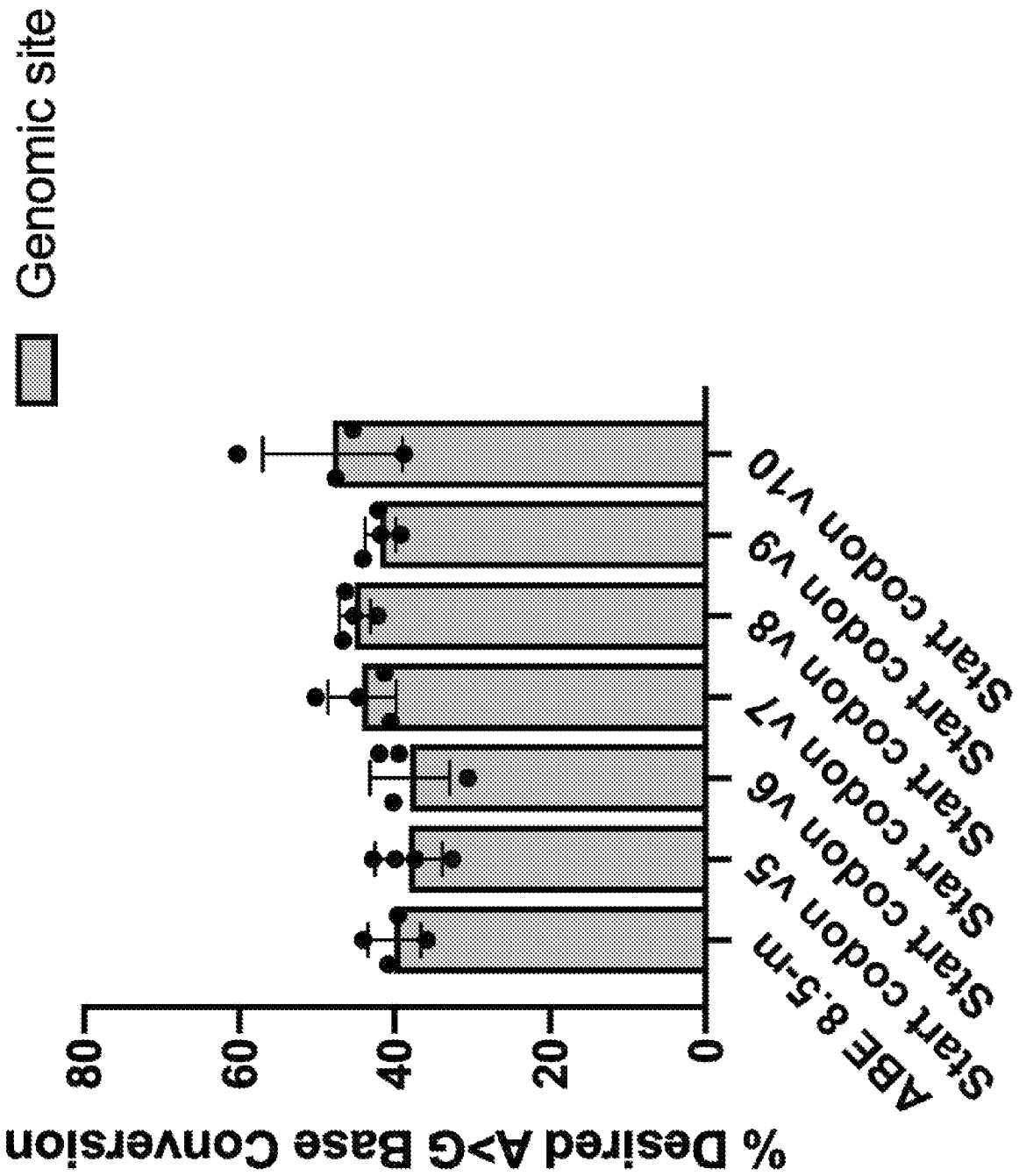

Genomic site
% Desired A>G Base Conversion
0
20
40
60
80
ABE 8.5-m
Start codon V5
Start codon V6
Start codon V7
Start codon V8
Start codon V9
Start codon V10

FIG. 2D

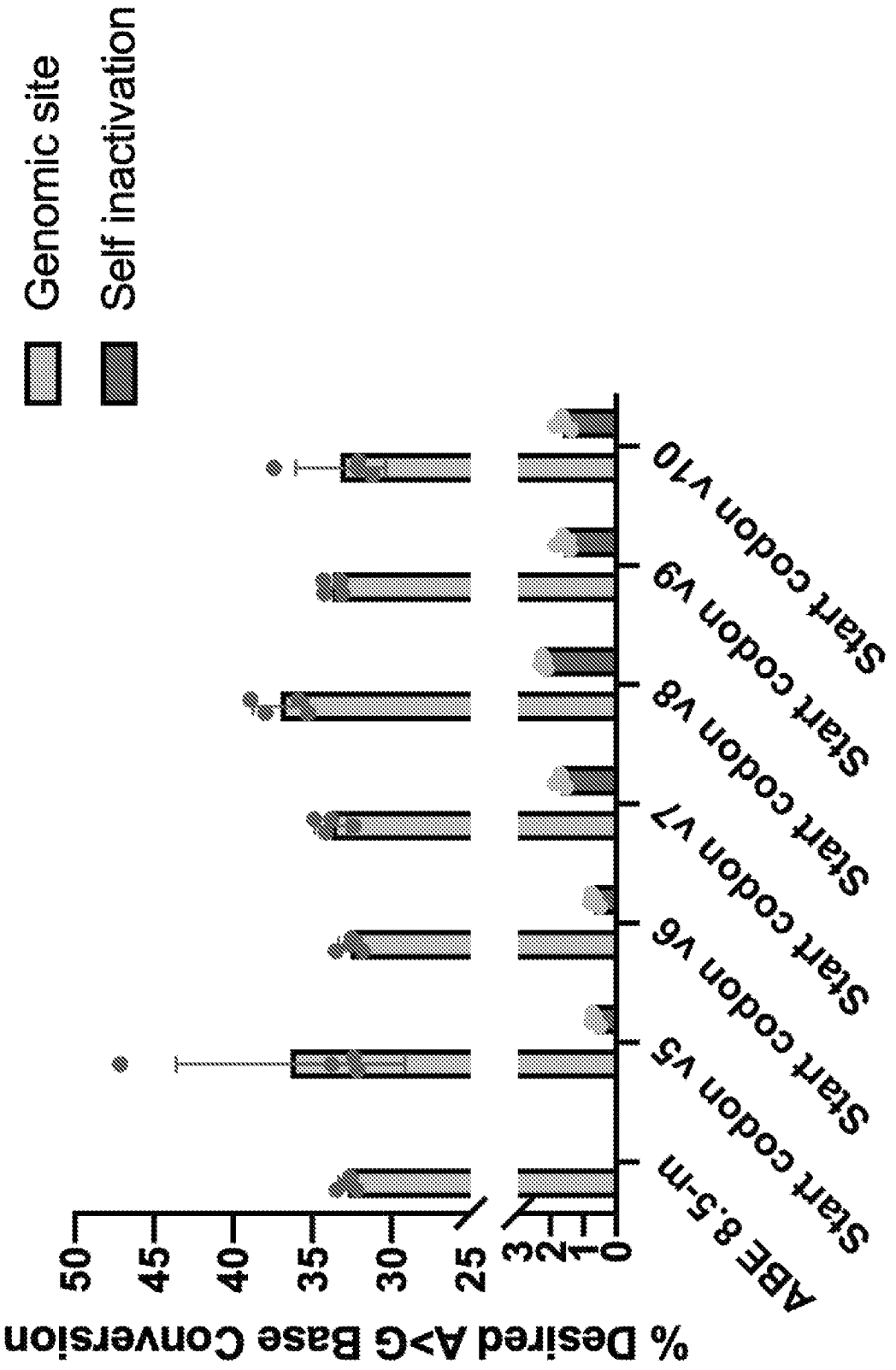

Genomic site
Self inactivation
% Desired A>G Base Conversion
50
45
40
35
30
25
3
2
1
0
ABE 8.5-m
Start codon V5
Start codon V6
Start codon V7
Start codon V8
Start codon V9
Start codon V10

FIG. 2E

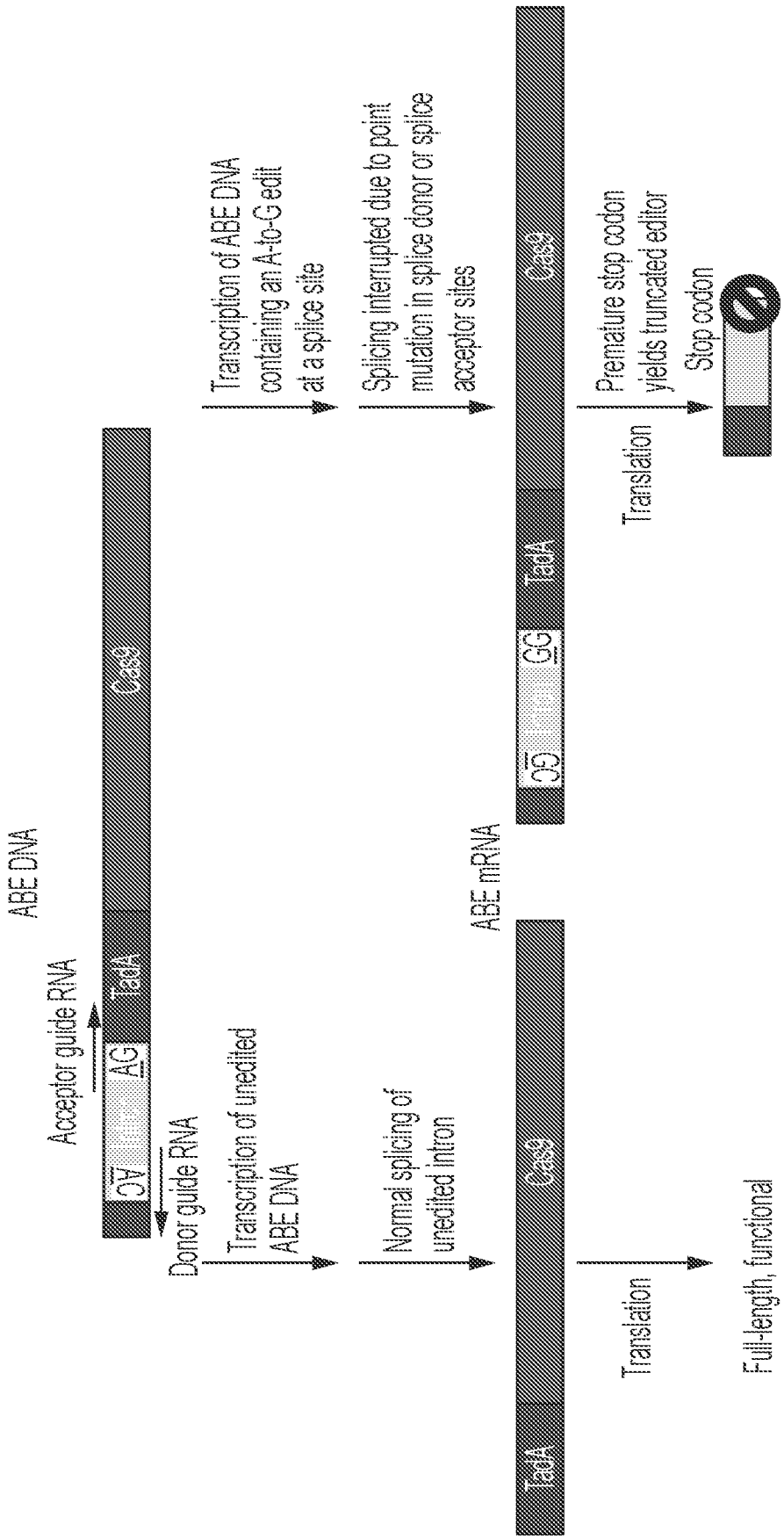

ABE DNA
Acceptor guide RNA
AG
TadA
Cas9
CT
Donor guide RNA
Transcription of unedited ABE DNA
Normal splicing of unedited intron
ABE mRNA
CT
GG
TadA
Cas9
Translation
Full-length, functional base editor enzyme
Transcription of ABE DNA containing an A-to-G edit at a splice site
Splicing interrupted due to point mutation in splice donor or splice acceptor sites
Premature stop codon yields truncated editor
Translation
Stop codon

FIG. 3A

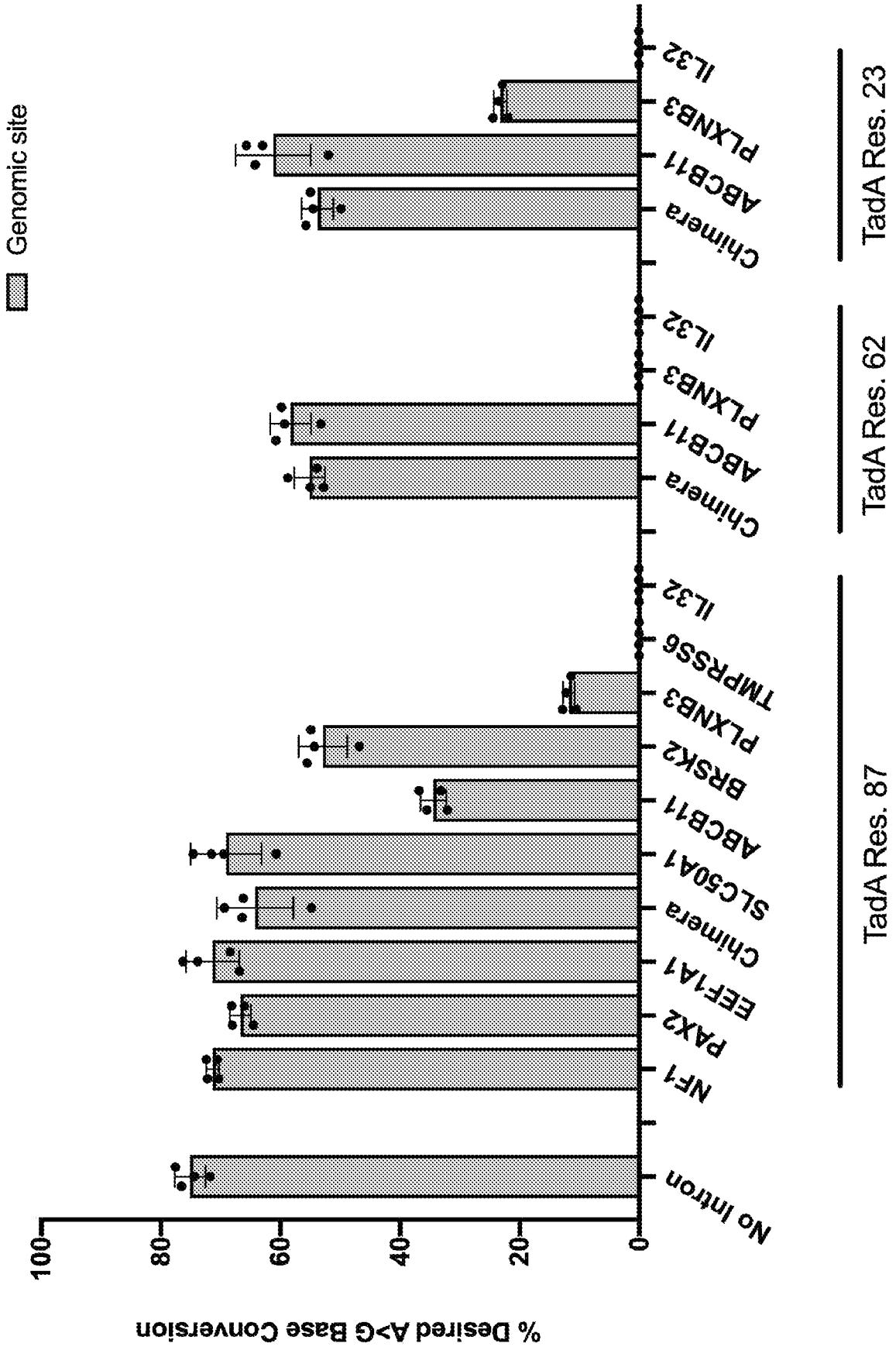

Genomic site
% Desired A>G Base Conversion
100
80
60
40
20
0
No Intron
NF1
PAX2
EEF1A1
Chimera
SLC50A1
ABCB11
BRSK2
PLXNB3
TMPRSS6
IL32
Chimera
ABCB11
PLXNB3
IL32
Chimera
ABCB11
PLXNB3
IL32
TadA Res. 87
TadA Res. 62
TadA Res. 23

FIG. 3B

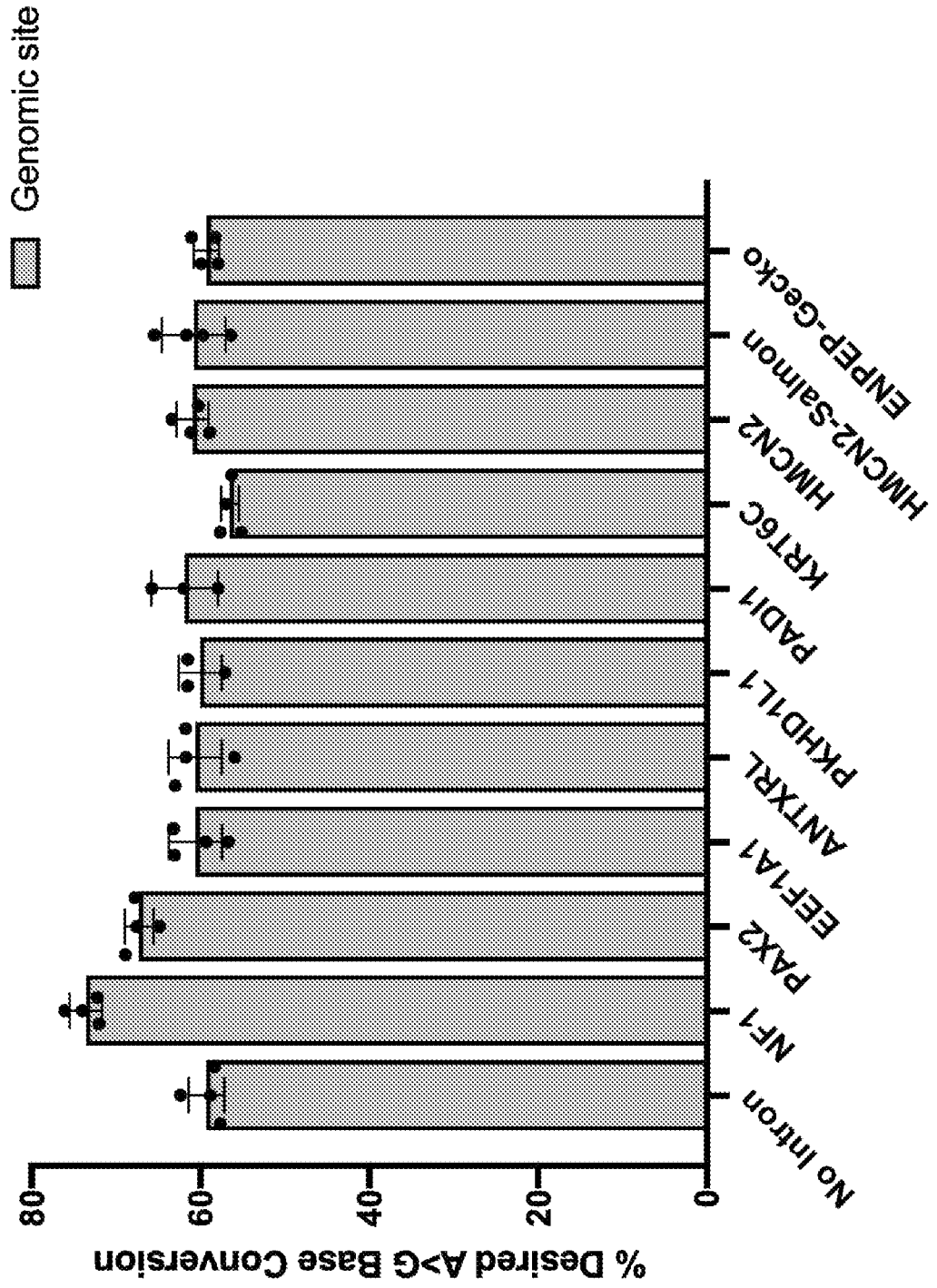

Genomic site
% Desired A>G Base Conversion
0
20
40
60
80
No Intron
NF1
PAX2
EEF1A1
ANTXRL
PKHD1L1
PADI1
KRT6C
HMCN2
HMCN2-Salmon
ENPEP-Gecko

FIG. 3C

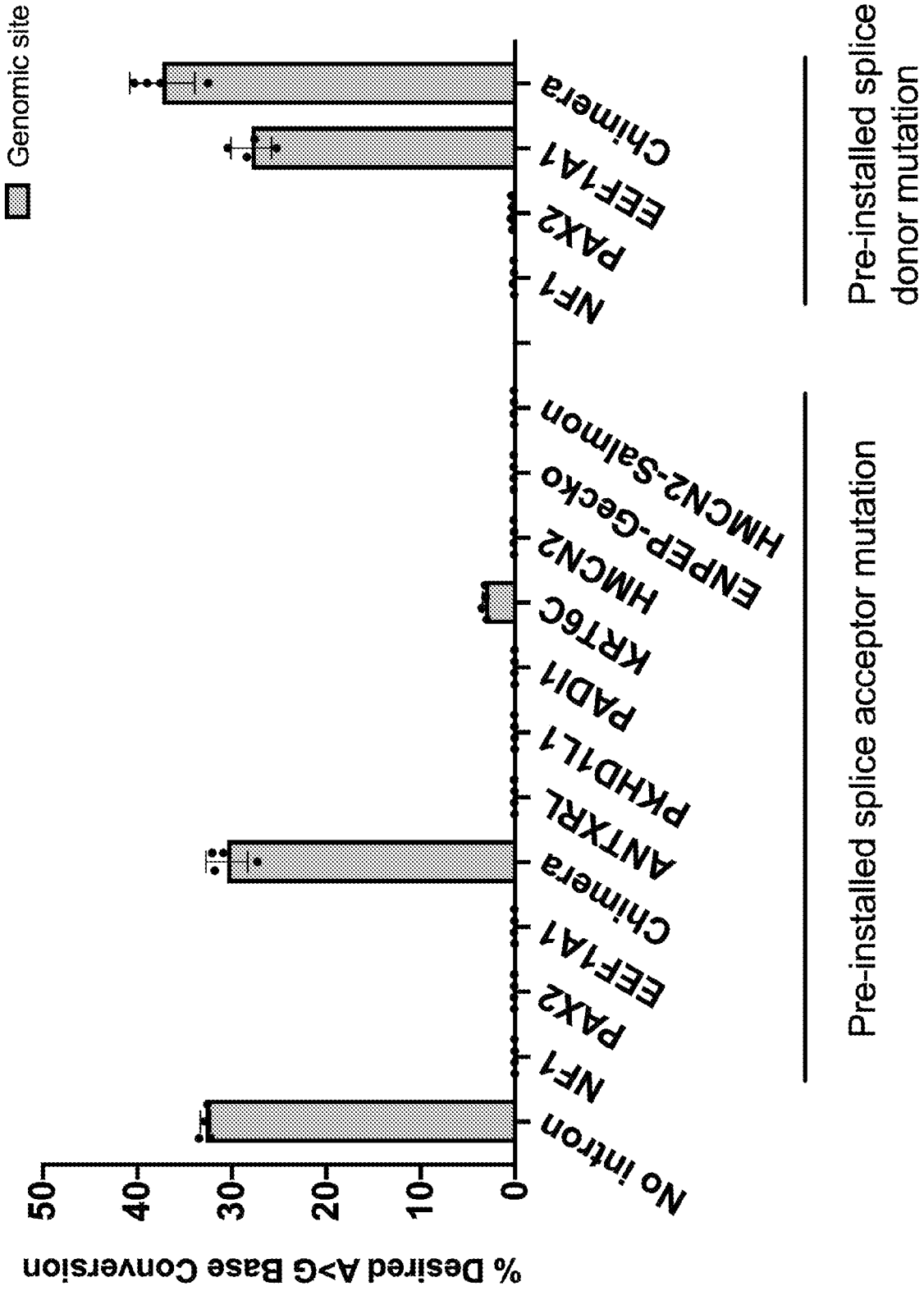

Genomic site
% Desired A>G Base Conversion
0
10
20
30
40
50
No intron
NF1
PAX2
EEF1A1
Chimera
ANTXRL
PKHD1L1
PADI1
KRT6C
HMCN2
ENPEP-Gecko
HMCN2-Salmon
NF1
PAX2
EEF1A1
Chimera
Pre-installed splice acceptor mutation
Pre-installed splice donor mutation

FIG. 3D

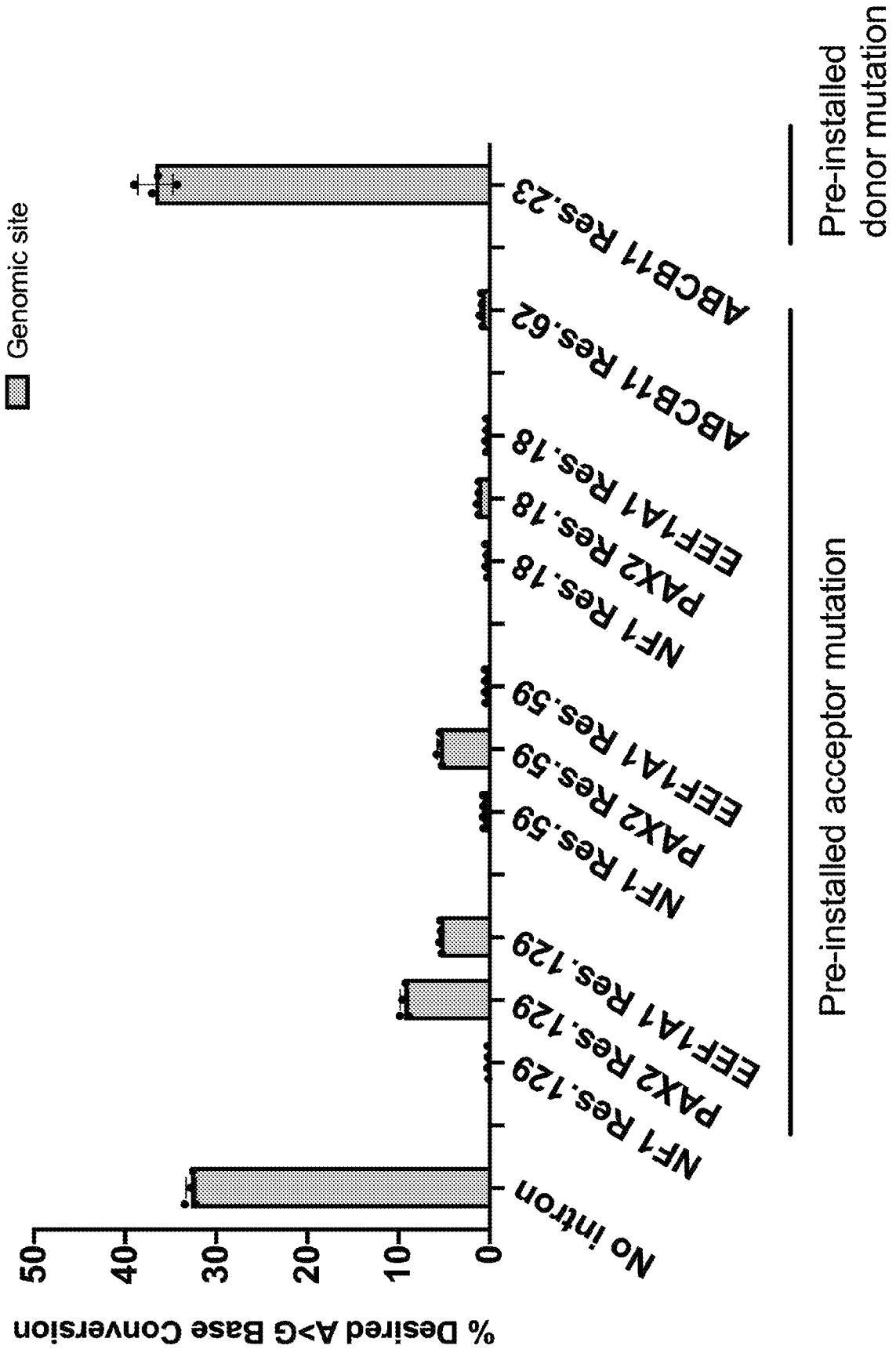

% Desired A>G Base Conversion
0
10
20
30
40
50
Genomic site
No intron
NF1 Res.129
PAX2 Res.129
EEF1A1 Res.129
NF1 Res.59
PAX2 Res.59
EEF1A1 Res.59
NF1 Res.18
PAX2 Res.18
EEF1A1 Res.18
ABCB11 Res.62
ABCB11 Res.23
Pre-installed acceptor mutation
Pre-installed donor mutation

FIG. 3E

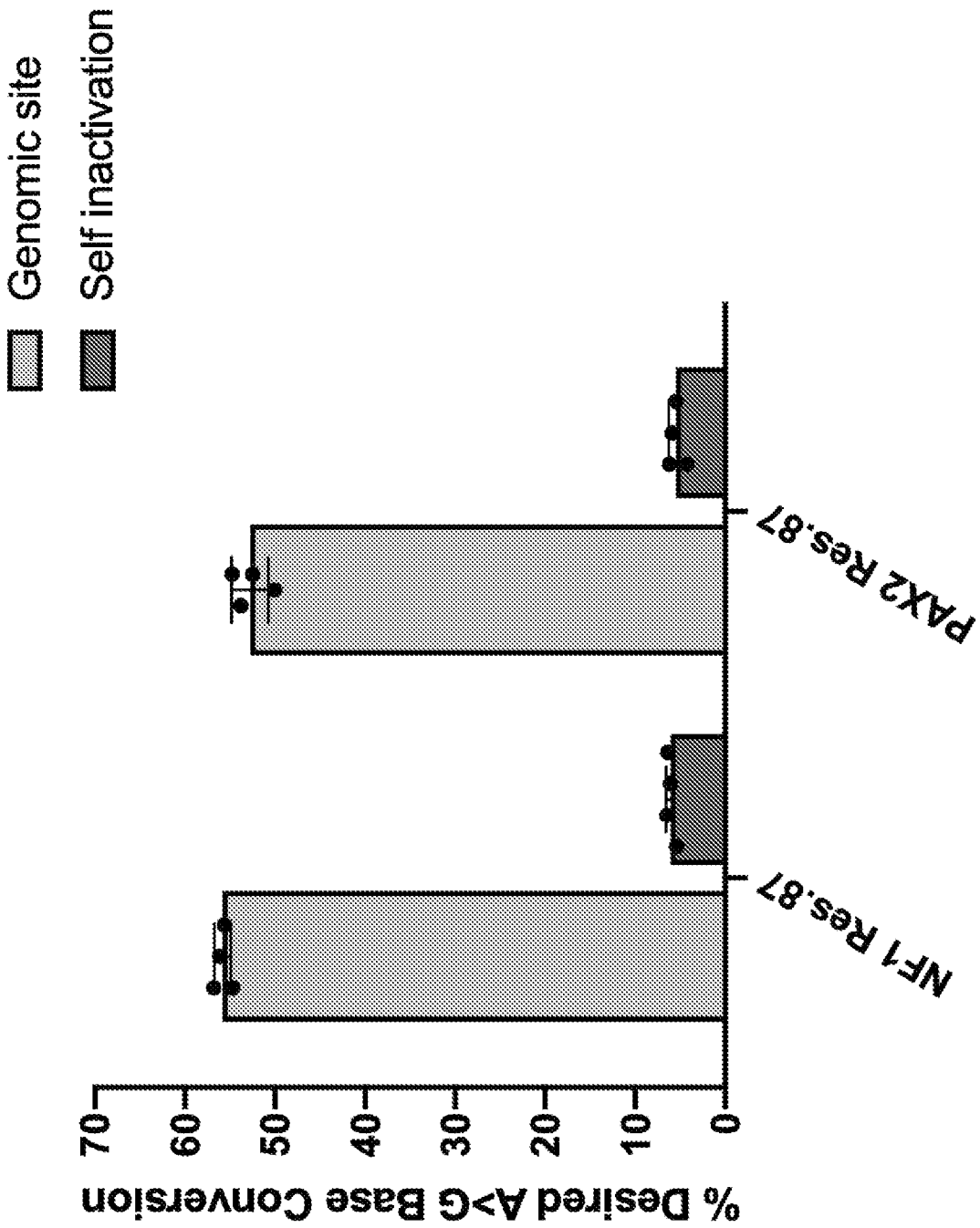

Genomic site
Self inactivation
% Desired A>G Base Conversion
0
10
20
30
40
50
60
70
NF1 Res.87
PAX2 Res.87

FIG. 3F

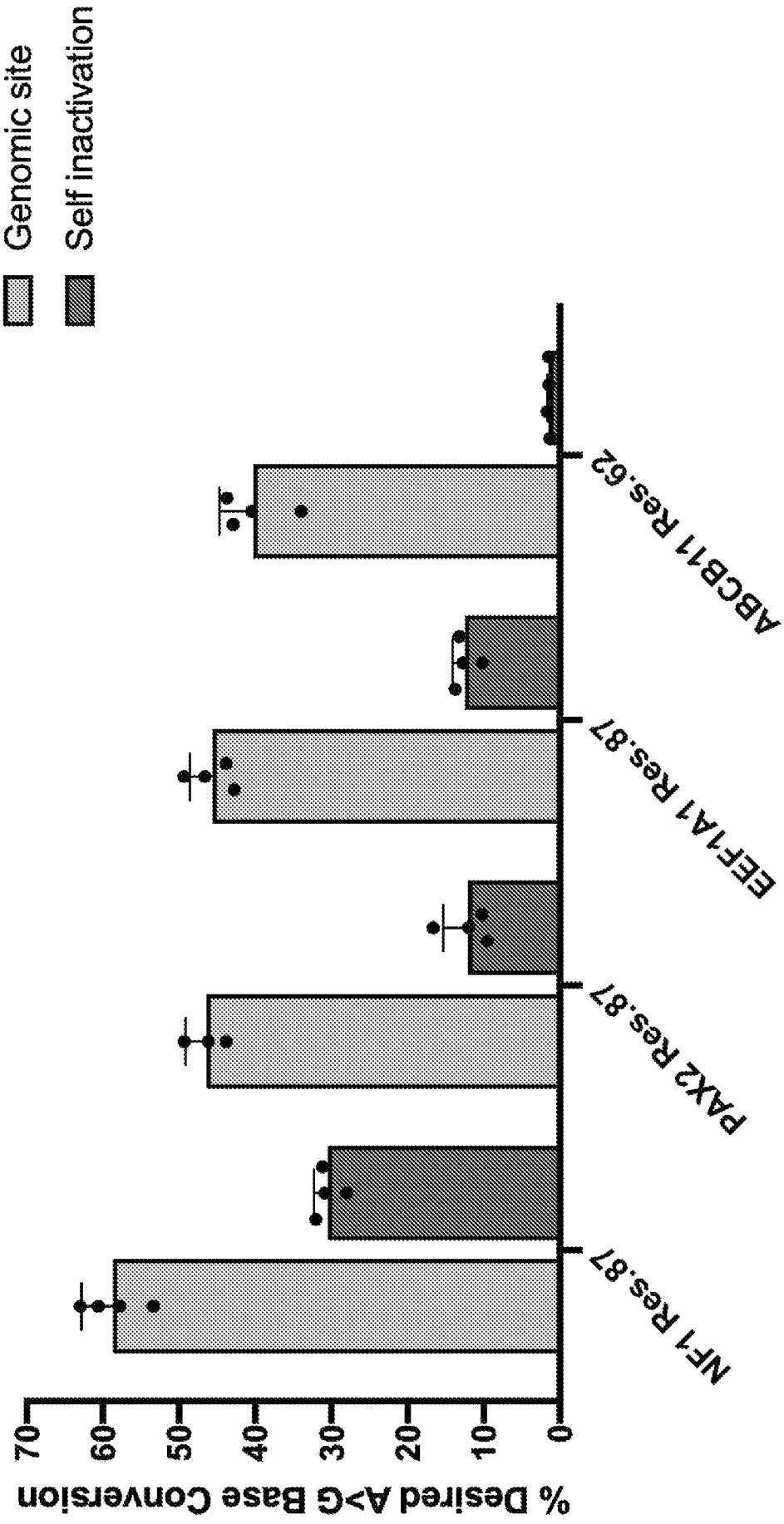

Genomic site
Self inactivation
% Desired A>G Base Conversion
0
10
20
30
40
50
60
70
NF1 Res.87
PAX2 Res.87
EEF1A1 Res.87
ABCB11 Res.62

FIG. 3G

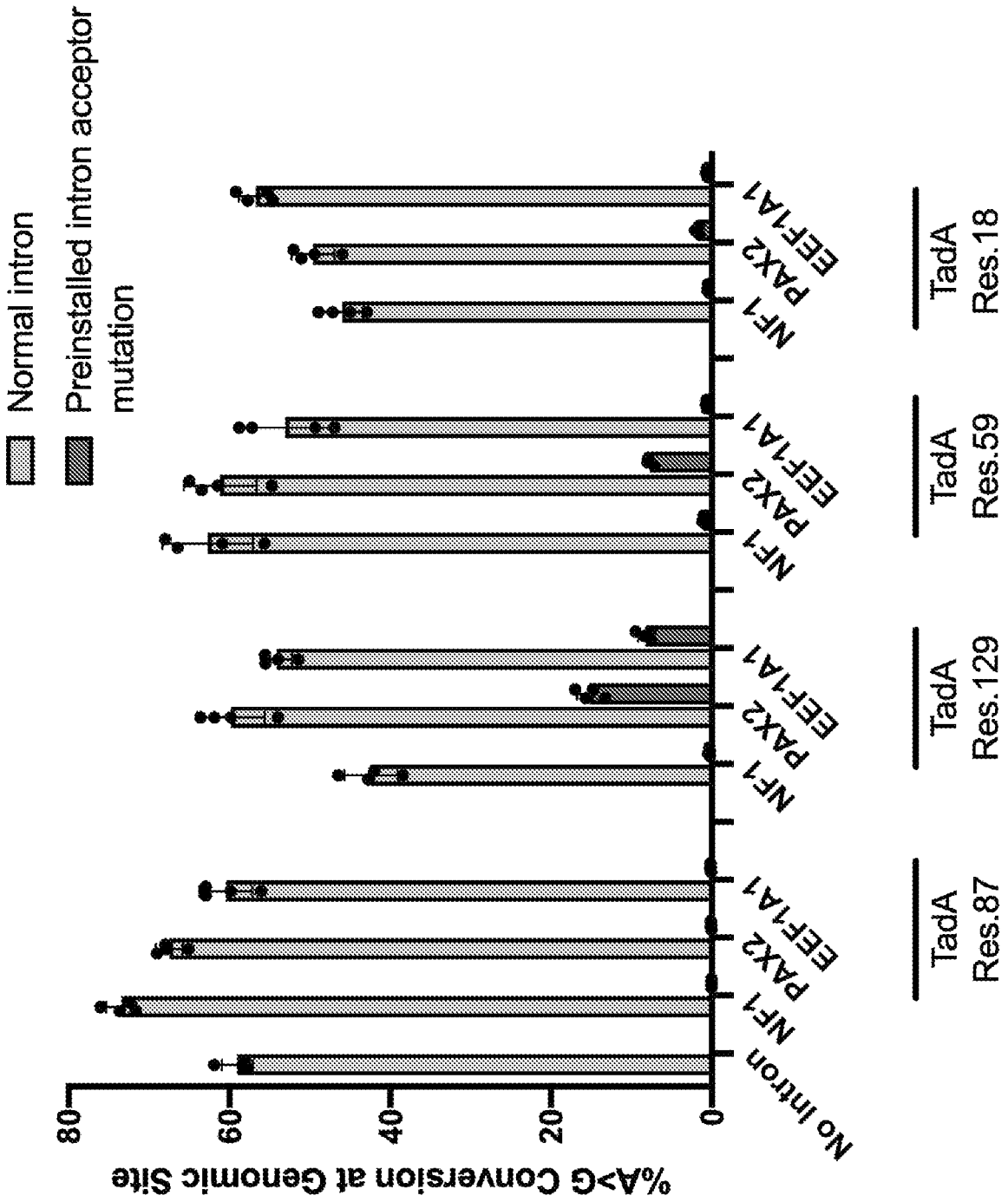

Normal intron
Preinstalled intron acceptor mutation
%A>G Conversion at Genomic Site
0
20
40
60
80
No intron
NF1
PAX2
EEF1A1
TadA Res.87
TadA Res.129
TadA Res.59
TadA Res.18

FIG. 3H

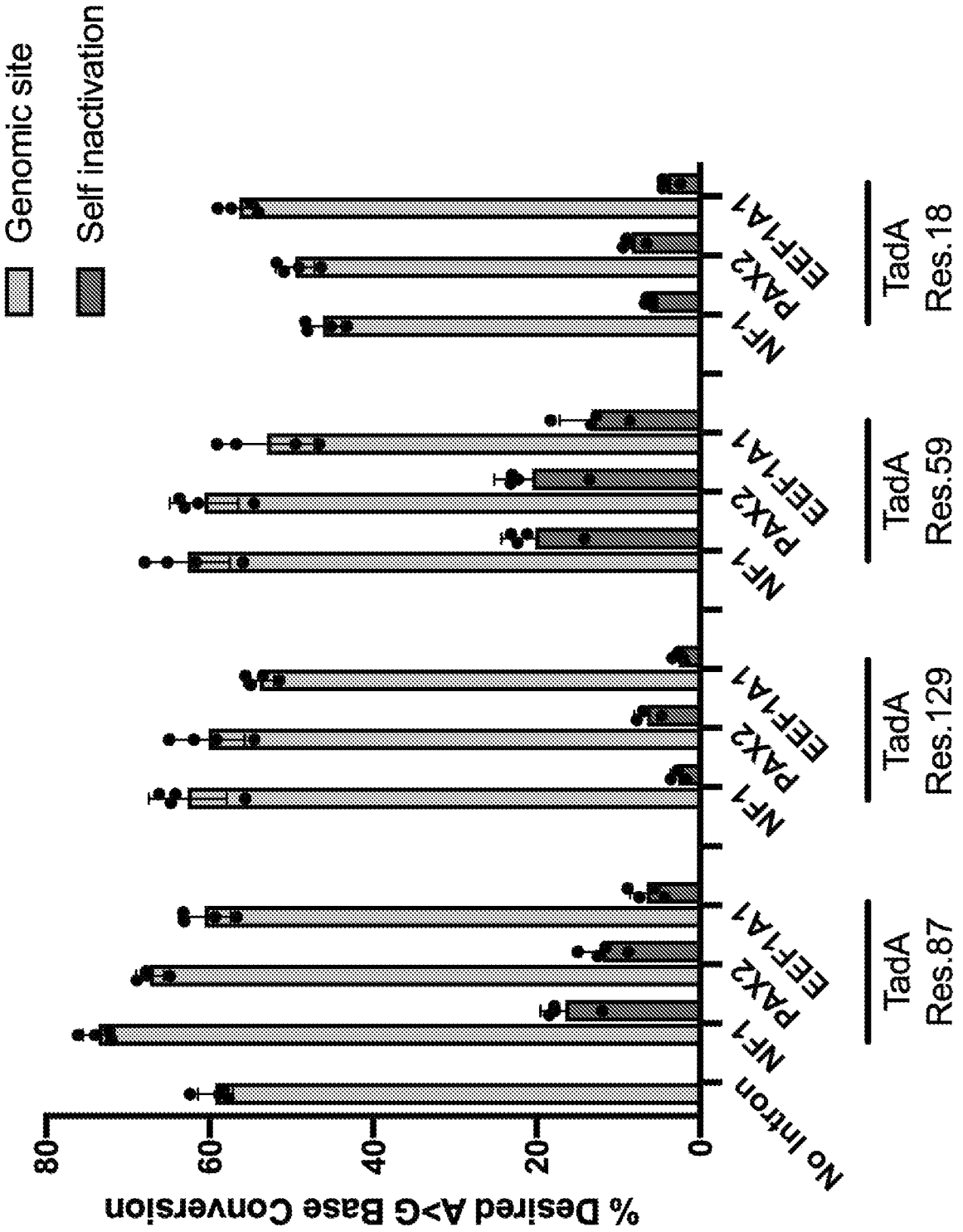

Genomic site
Self inactivation
% Desired A>G Base Conversion
0
20
40
60
80
No intron
NF1
PAX2
EEF1A1
TadA Res.87
TadA Res.129
TadA Res.59
TadA Res.18

FIG. 3I

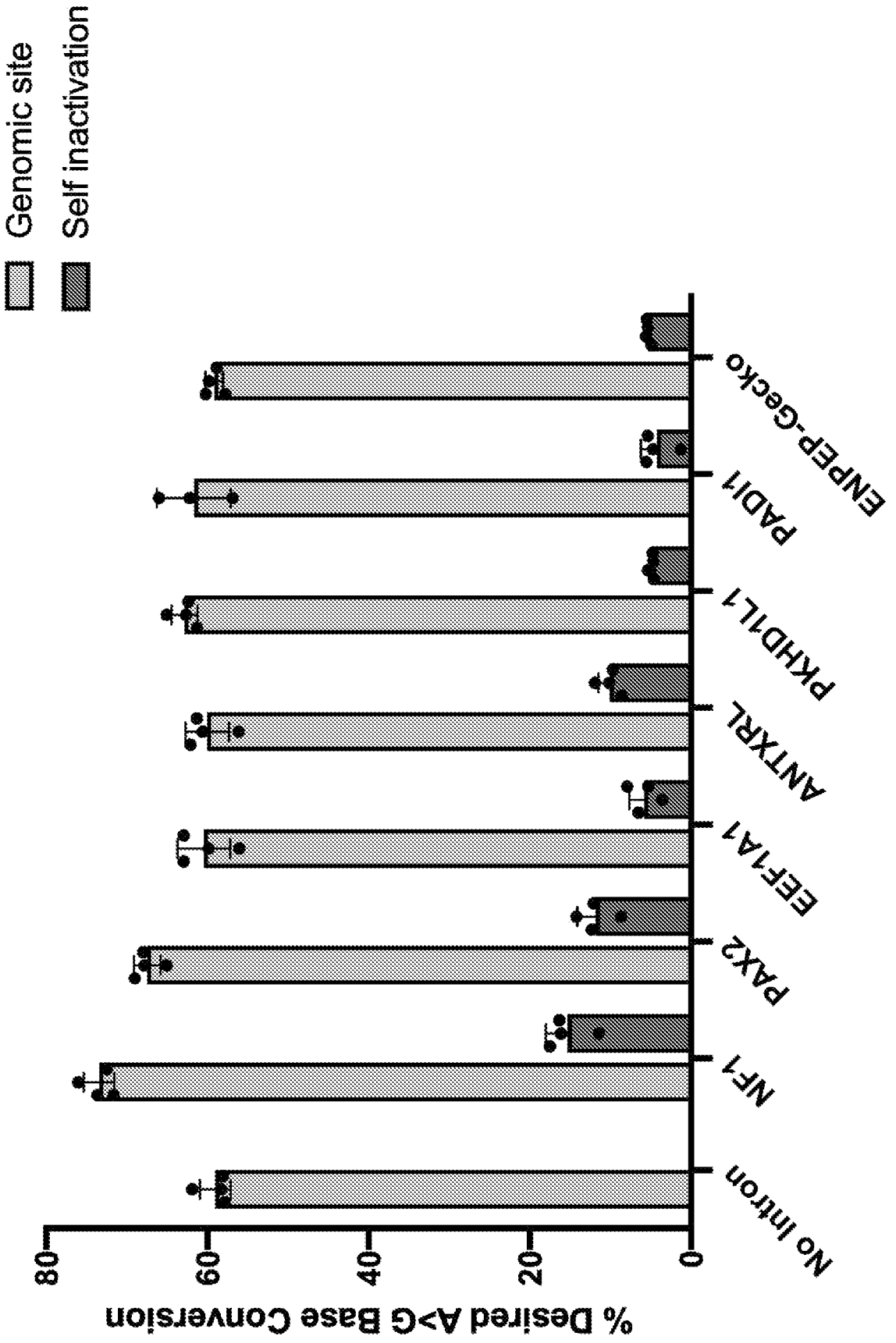

Genomic site
Self inactivation
% Desired A>G Base Conversion
0
20
40
60
80
No Intron
NF1
PAX2
EEF1A1
ANTXRL
PKHD1L1
PADI1
ENPEP-Gecko

FIG. 3J

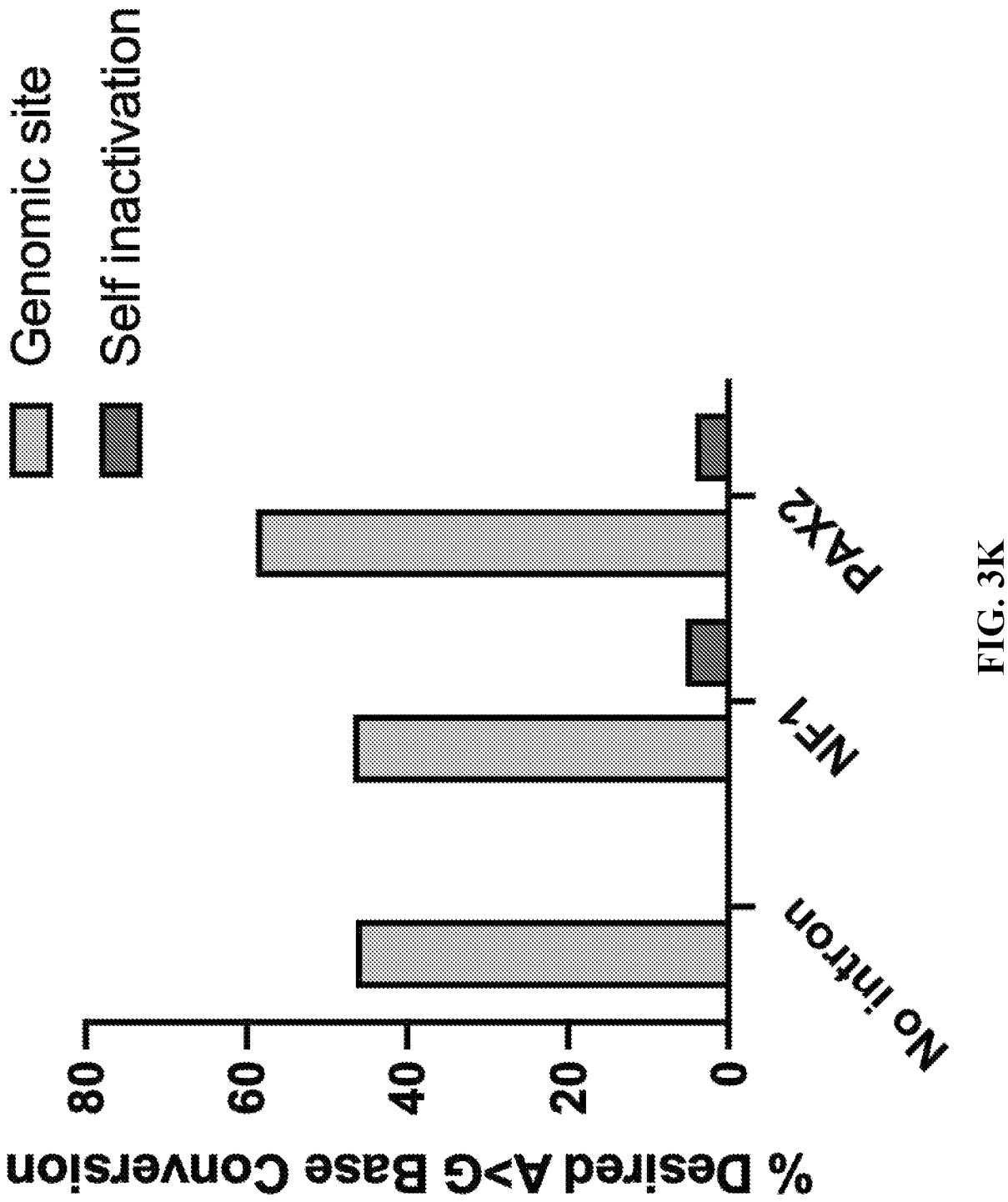

Genomic site
Self inactivation
% Desired A>G Base Conversion
0
20
40
60
80
No intron
NF-1
PAX2

FIG. 3K


NF1

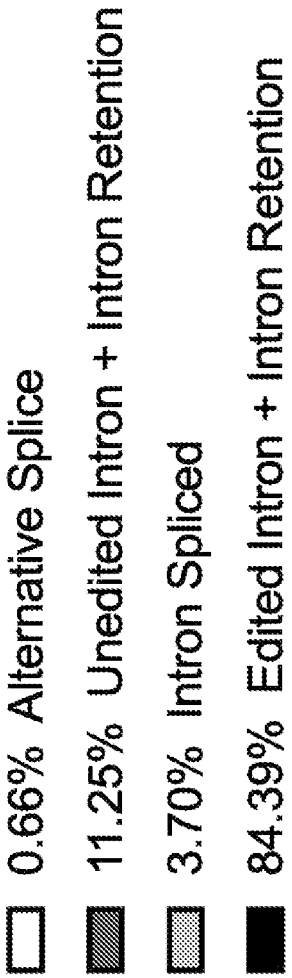

0.66% Alternative Splice
11.25% Unedited Intron + Intron Retention
3.70% Intron Spliced
84.39% Edited Intron + Intron Retention

FIG. 3L


1.91% Alternative Splice
2.27% Unedited Intron + Intron Retention
16.52% Intron Spliced
79.30% Edited Intron + Intron Retention
PAX2

FIG. 3M

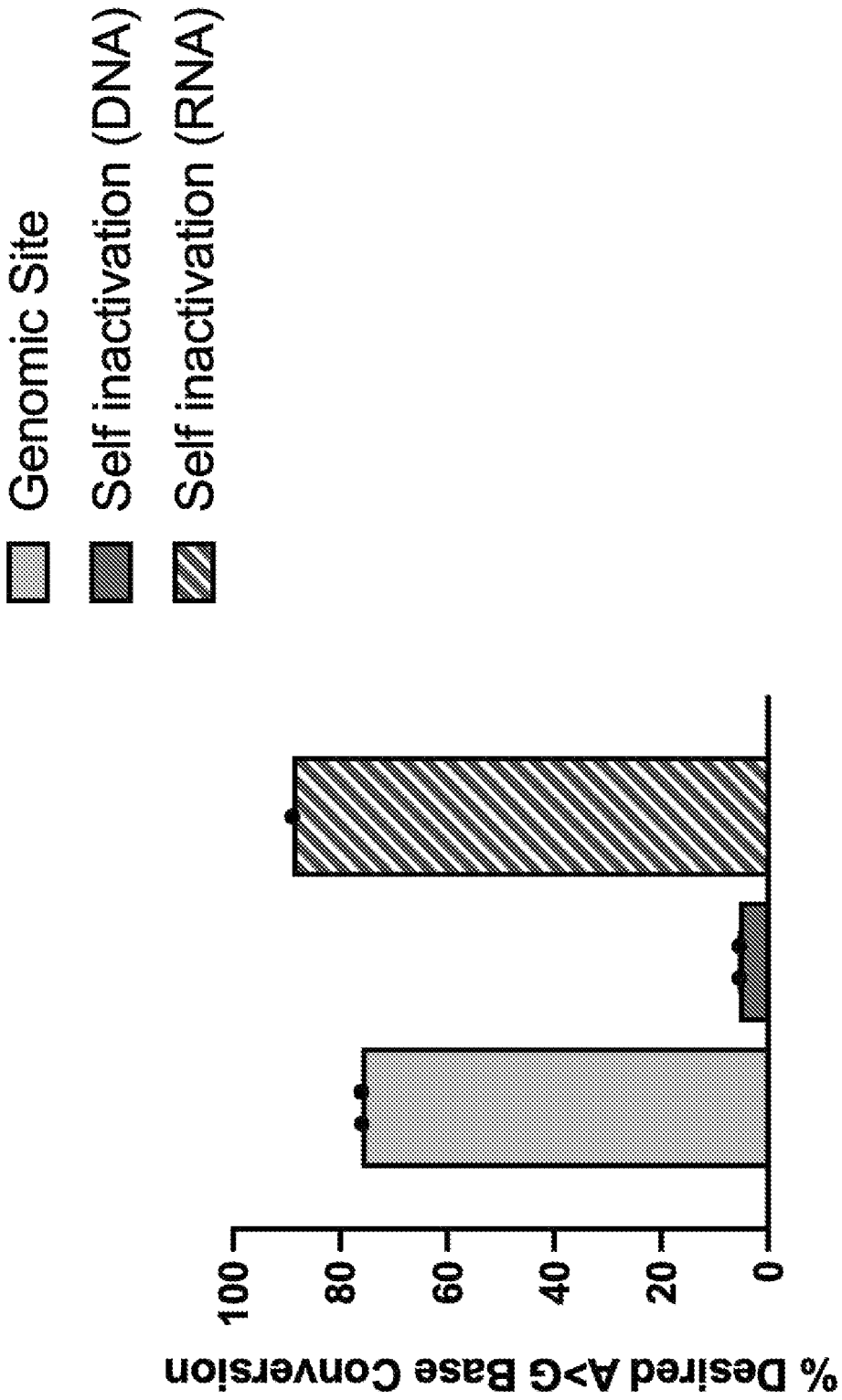

Genomic Site
Self inactivation (DNA)
Self inactivation (RNA)
% Desired A>G Base Conversion
0
20
40
60
80
100

FIG. 3N

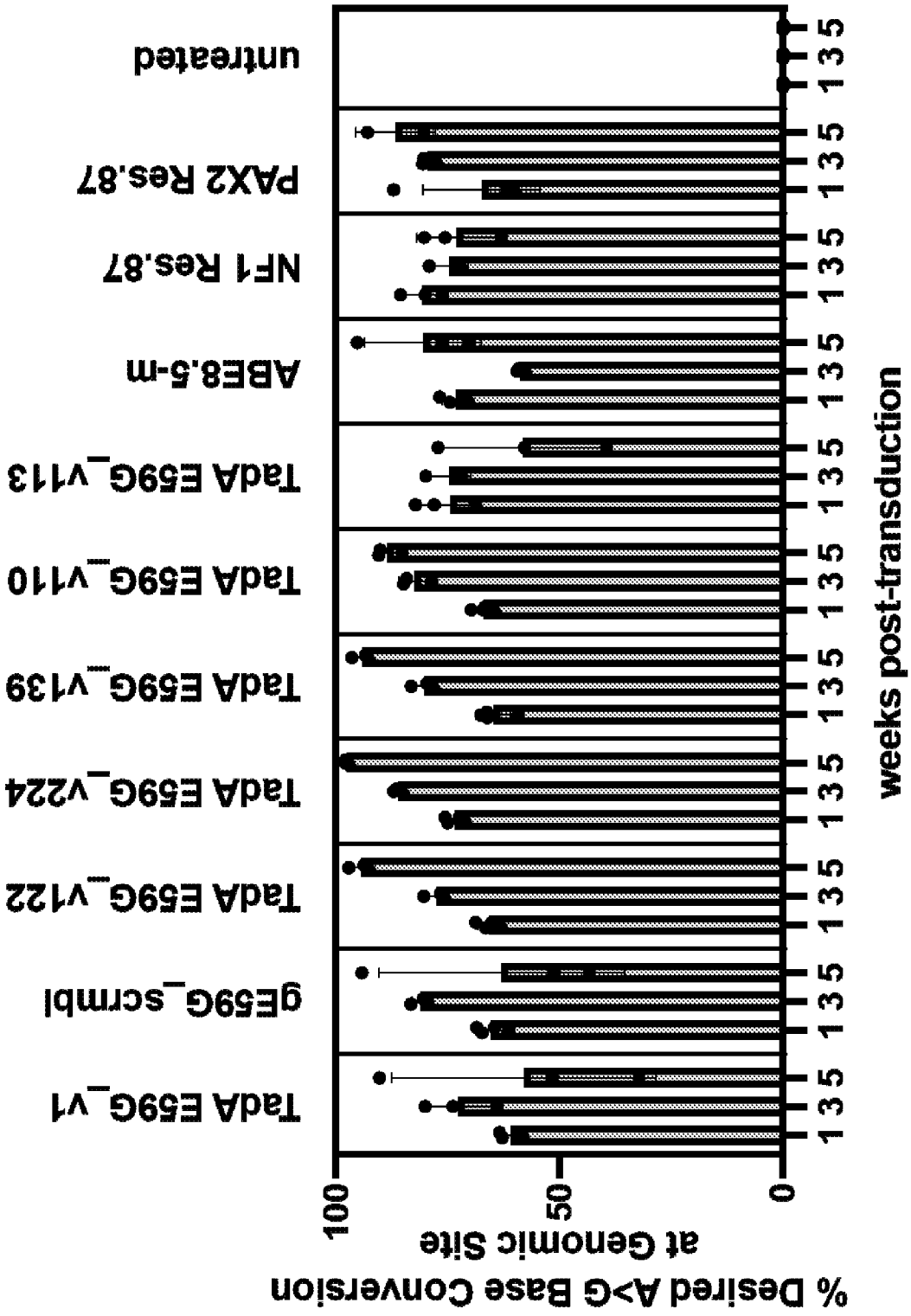

% Desired A>G Base Conversion at Genomic Site
100
50
0
weeks post-transduction
1 3 5
TadA E59G_V1
gE59G_scrmbl
TadA E59G_V122
TadA E59G_V224
TadA E59G_V139
TadA E59G_V110
TadA E59G_V113
ABE8.5-m
NF1 Res.87
PAX2 Res.87
untreated

FIG. 4A

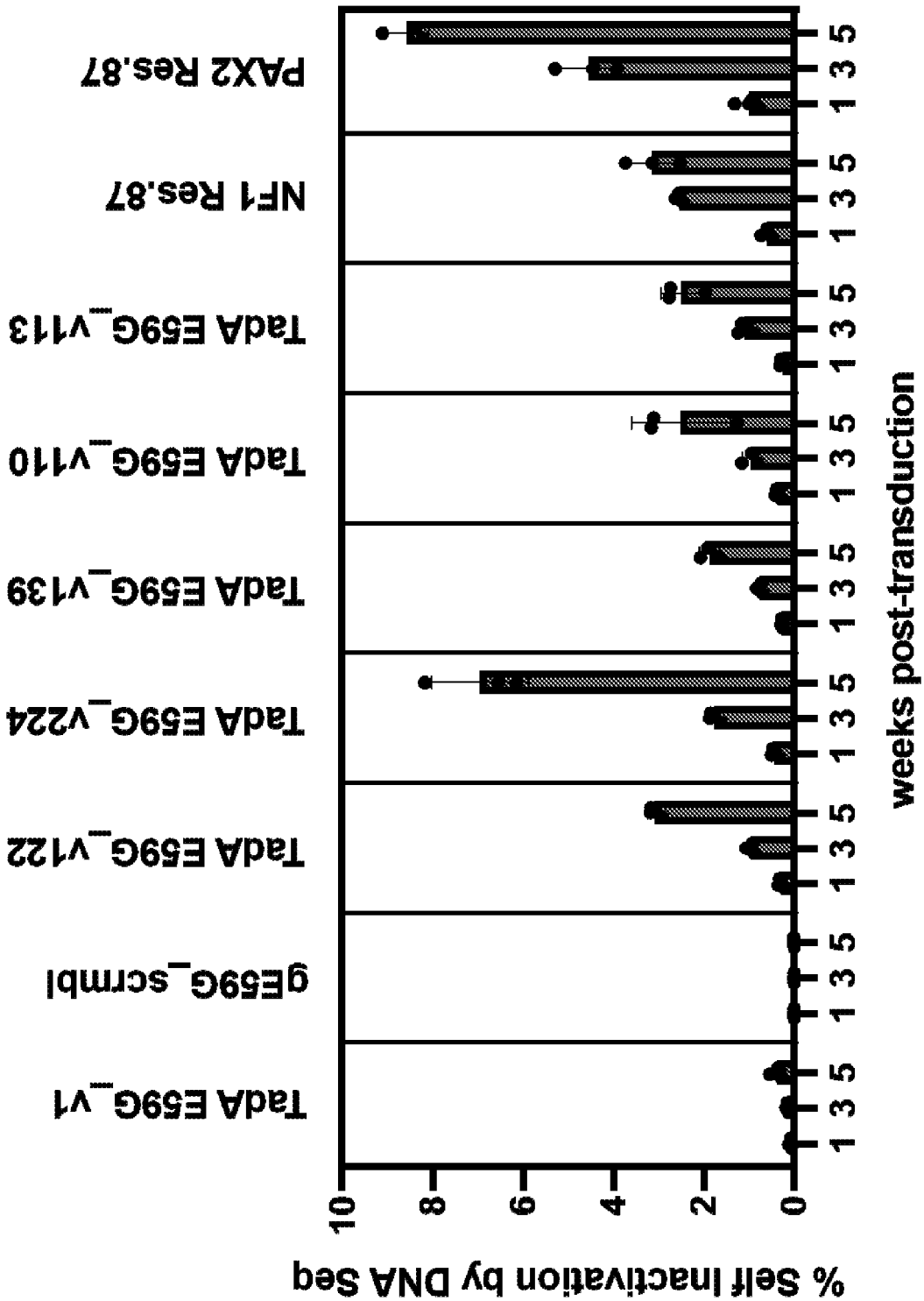

% Self Inactivation by DNA Seq
0
2
4
6
8
10
weeks post-transduction
1 3 5
TadA E59G_v1
gE59G_scrmbl
TadA E59G_v122
TadA E59G_v224
TadA E59G_v139
TadA E59G_v110
TadA E59G_v113
NF1 Res.87
PAX2 Res.87

FIG. 4B

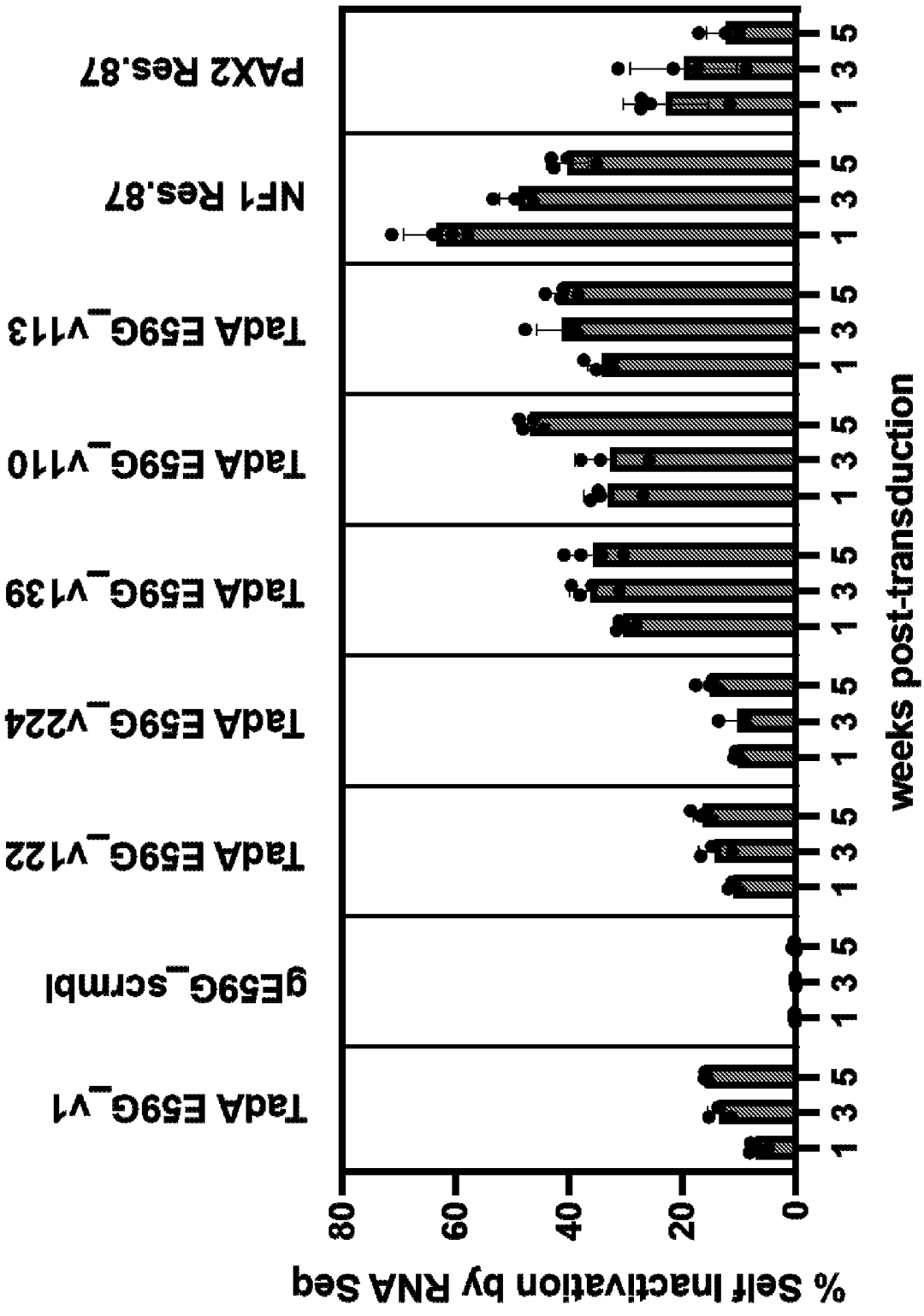

% Self Inactivation by RNA Seq
0
20
40
60
80
weeks post-transduction
1 3 5
TadA E59G_v1
gE59G_scrmbl
TadA E59G_v122
TadA E59G_v224
TadA E59G_v139
TadA E59G_v110
TadA E59G_v113
NF1 Res.87
PAX2 Res.87

FIG. 4C

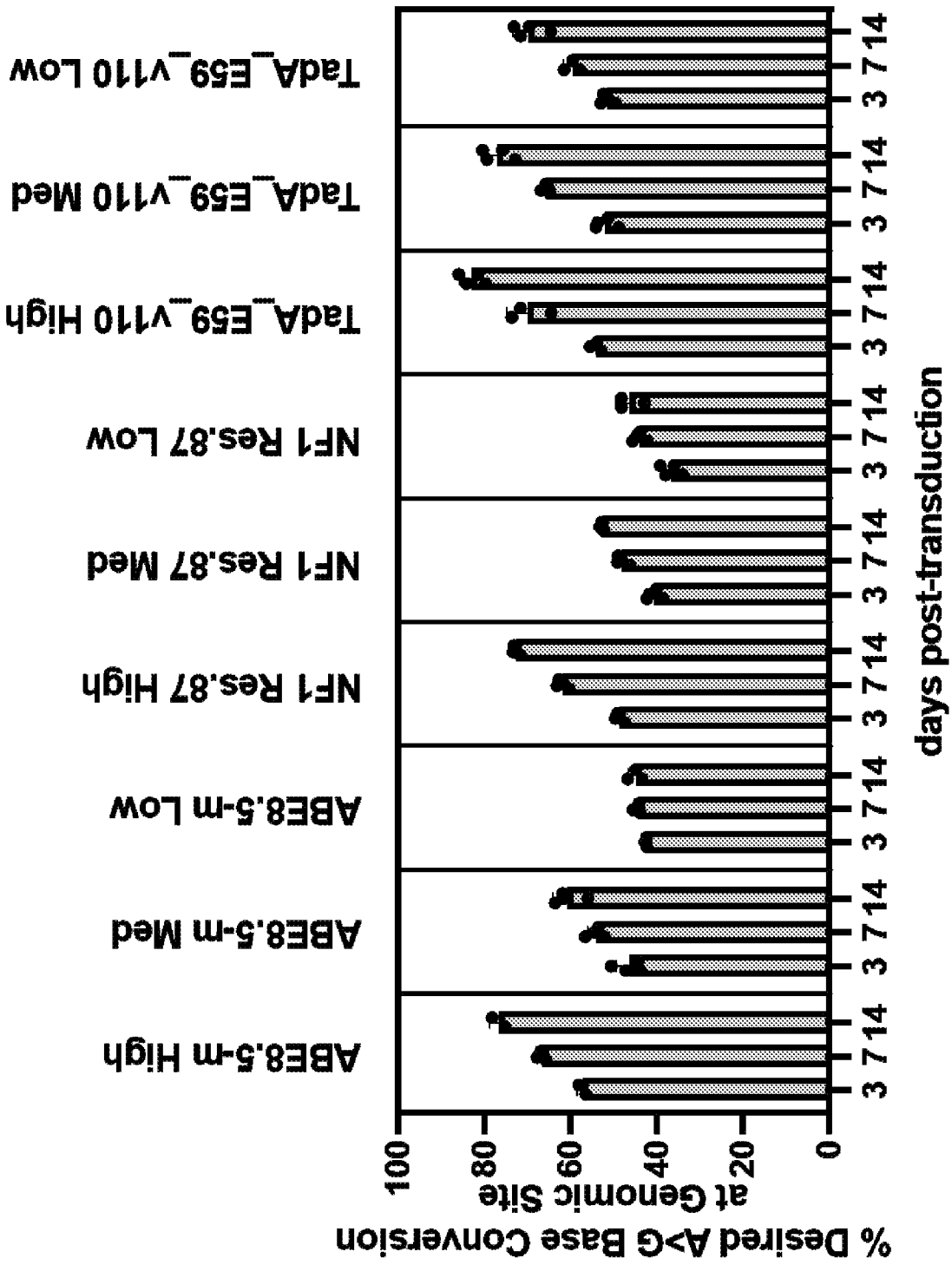

% Desired A>G Base Conversion at Genomic Site
0
20
40
60
80
100
3 7 14
days post-transduction
ABE8.5-m High
ABE8.5-m Med
ABE8.5-m Low
NF1 Res.87 High
NF1 Res.87 Med
NF1 Res.87 Low
TadA_E59_V110 High
TadA_E59_V110 Med
TadA_E59_V110 Low

FIG. 5A

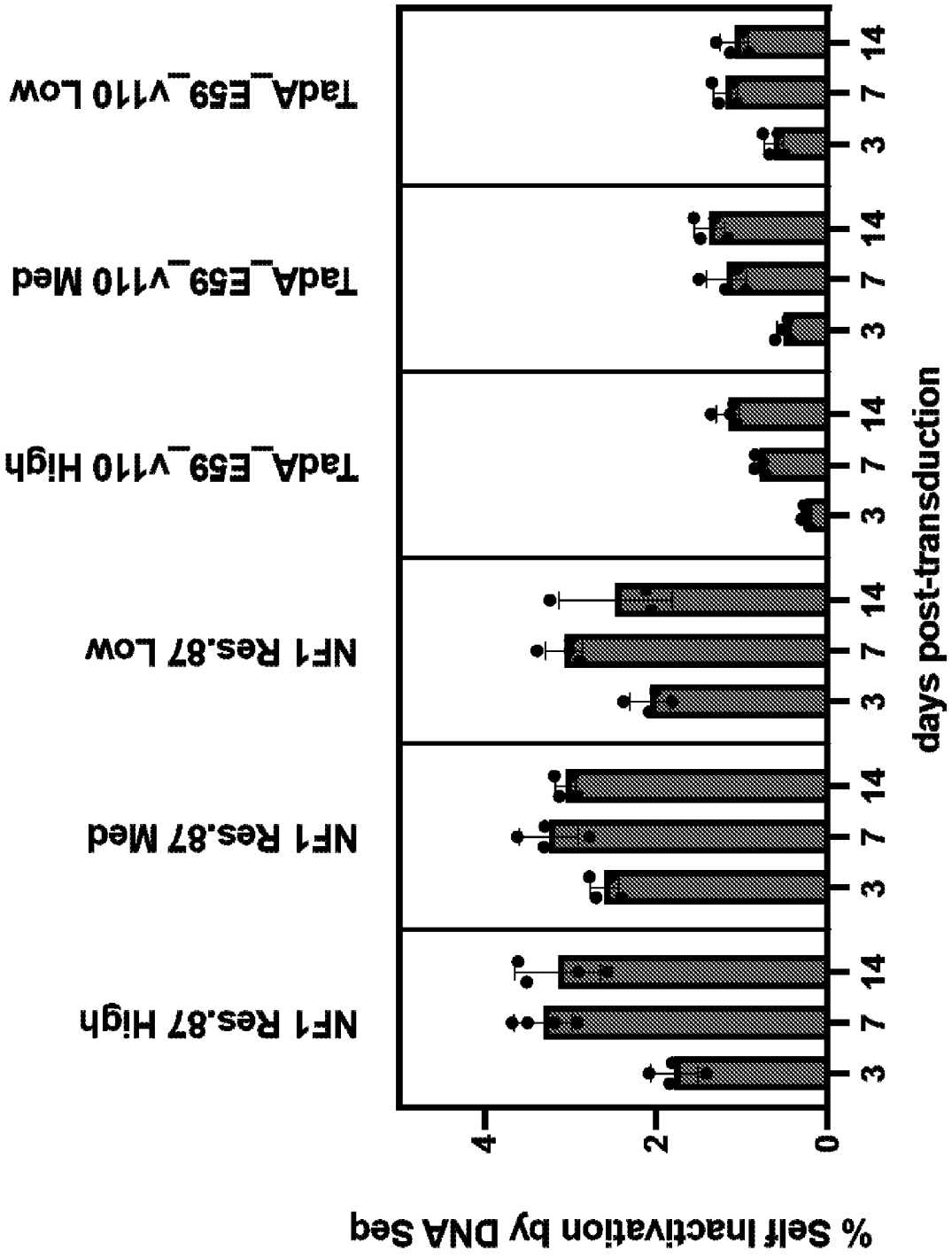

% Self Inactivation by DNA Seq
0
2
4
days post-transduction
3
7
14
NF1 Res.87 High
NF1 Res.87 Med
NF1 Res.87 Low
TadA_E59_V110 High
TadA_E59_V110 Med
TadA_E59_V110 Low

FIG. 5B

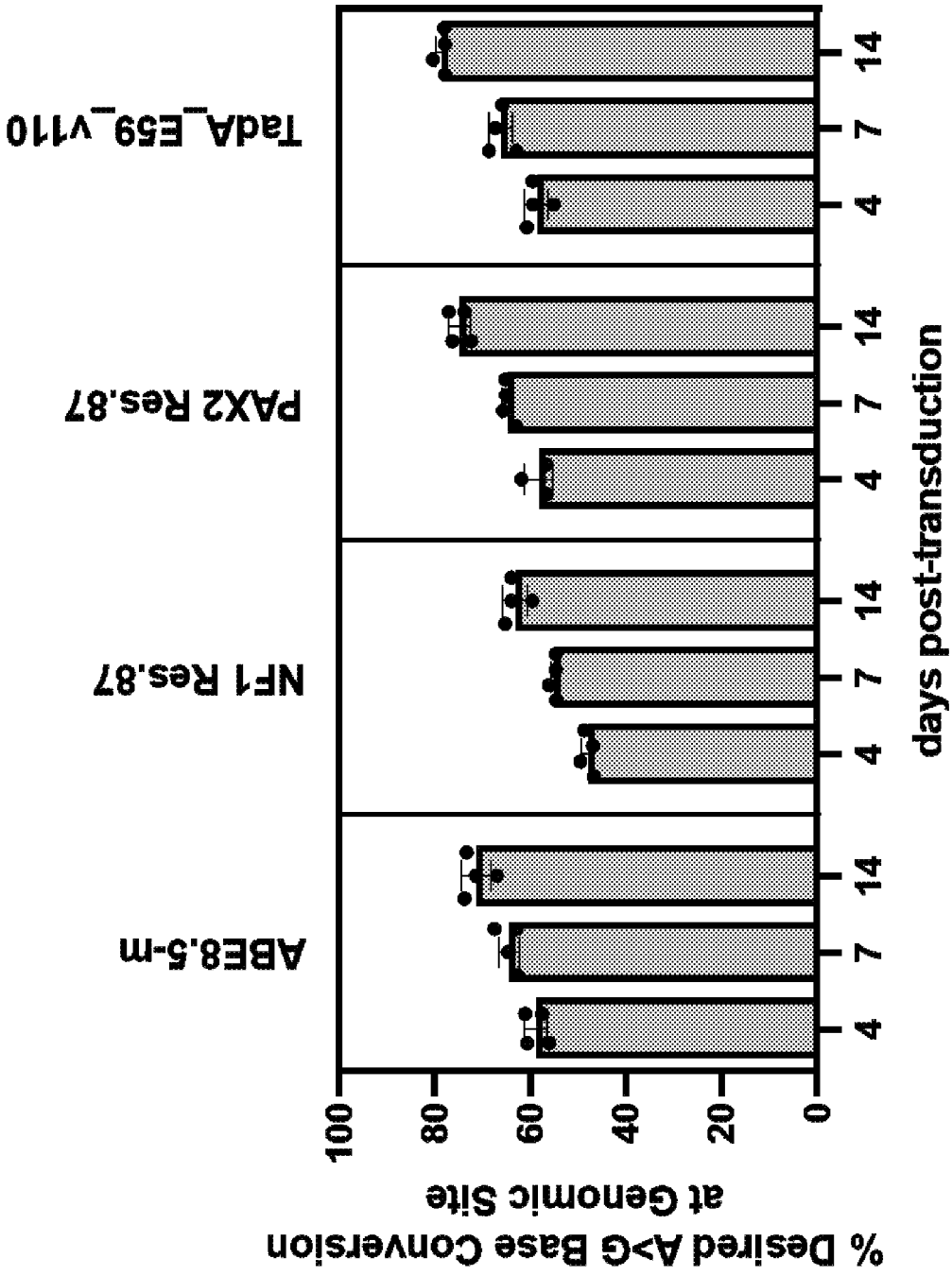

% Desired A>G Base Conversion
at Genomic Site
100
80
60
40
20
0
4
7
14
days post-transduction
ABE8.5-m
NF1 Res.87
PAX2 Res.87
TadA_E59_V110

FIG. 6A

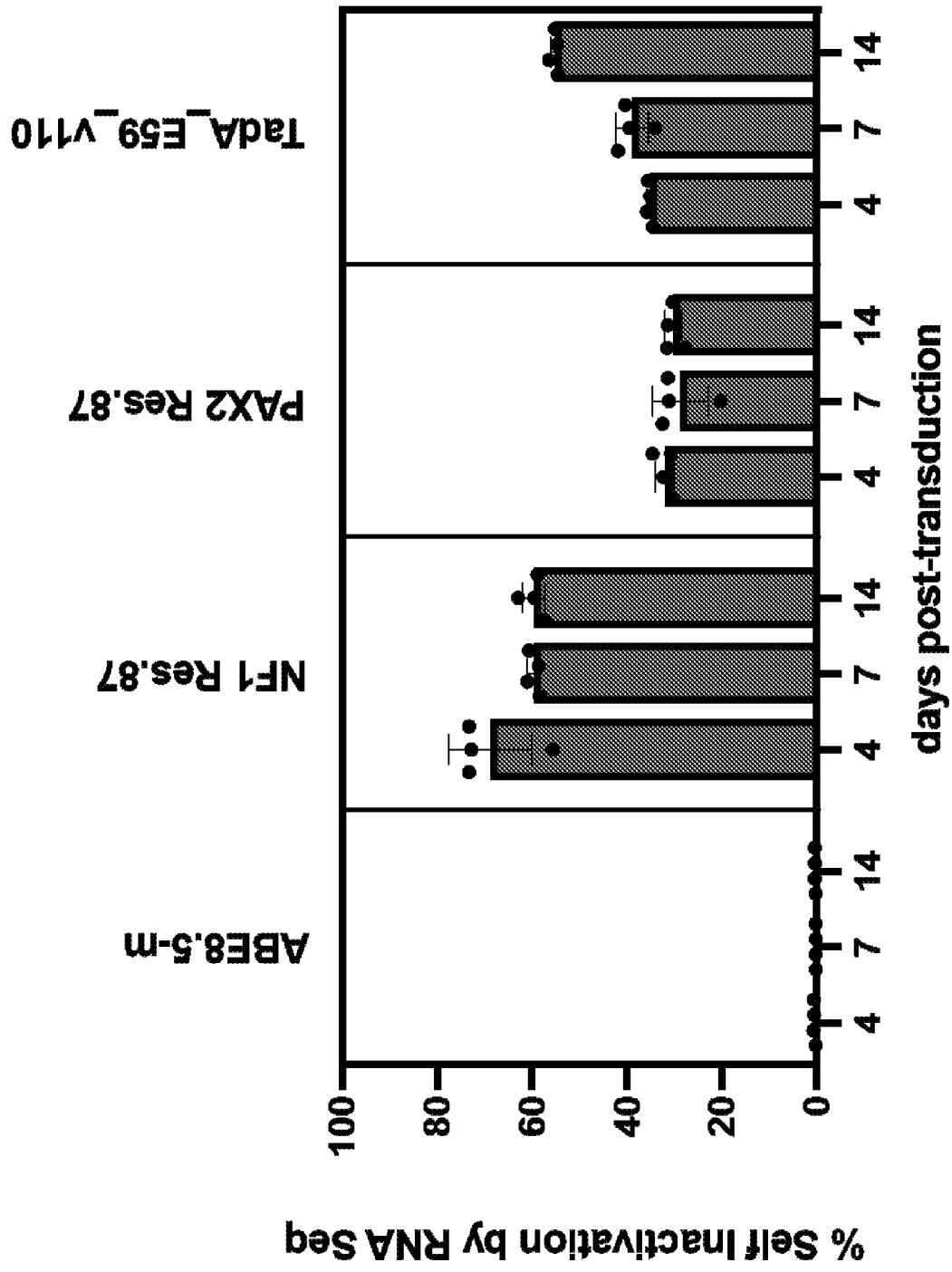

% Self Inactivation by RNA Seq
0
20
40
60
80
100
4
7
14
4
7
14
4
7
14
4
7
14
days post-transduction
ABE8.5-m
NF1 Res.87
PAX2 Res.87
TadA_E59_V110

FIG. 6B

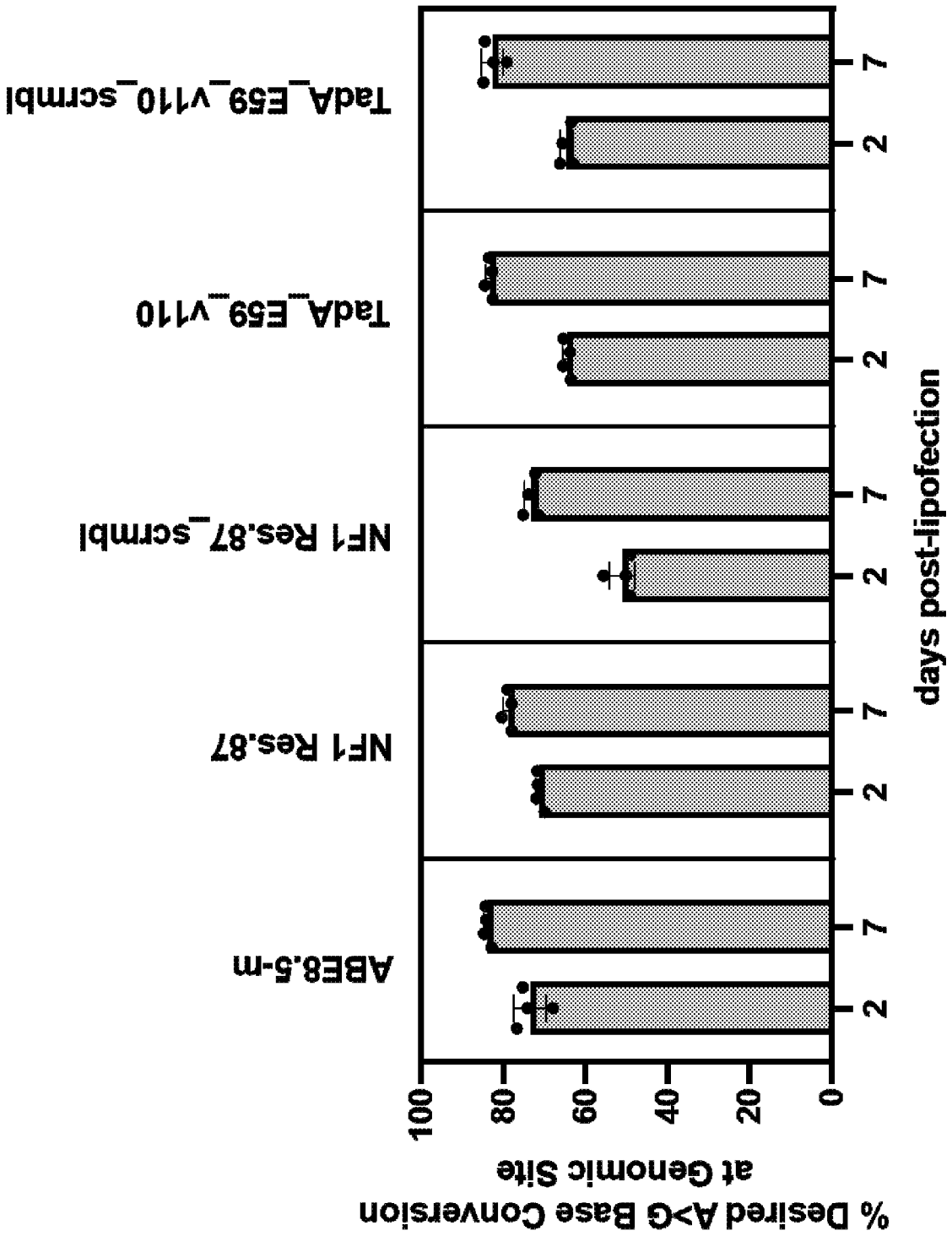

% Desired A>G Base Conversion at Genomic Site
100
80
60
40
20
0
ABE8.5-m
NF1 Res.87
NF1 Res.87_scrmbl
TadA_E59_V110
TadA_E59_V110_scrmbl
2
7
days post-lipofection

FIG. 7A

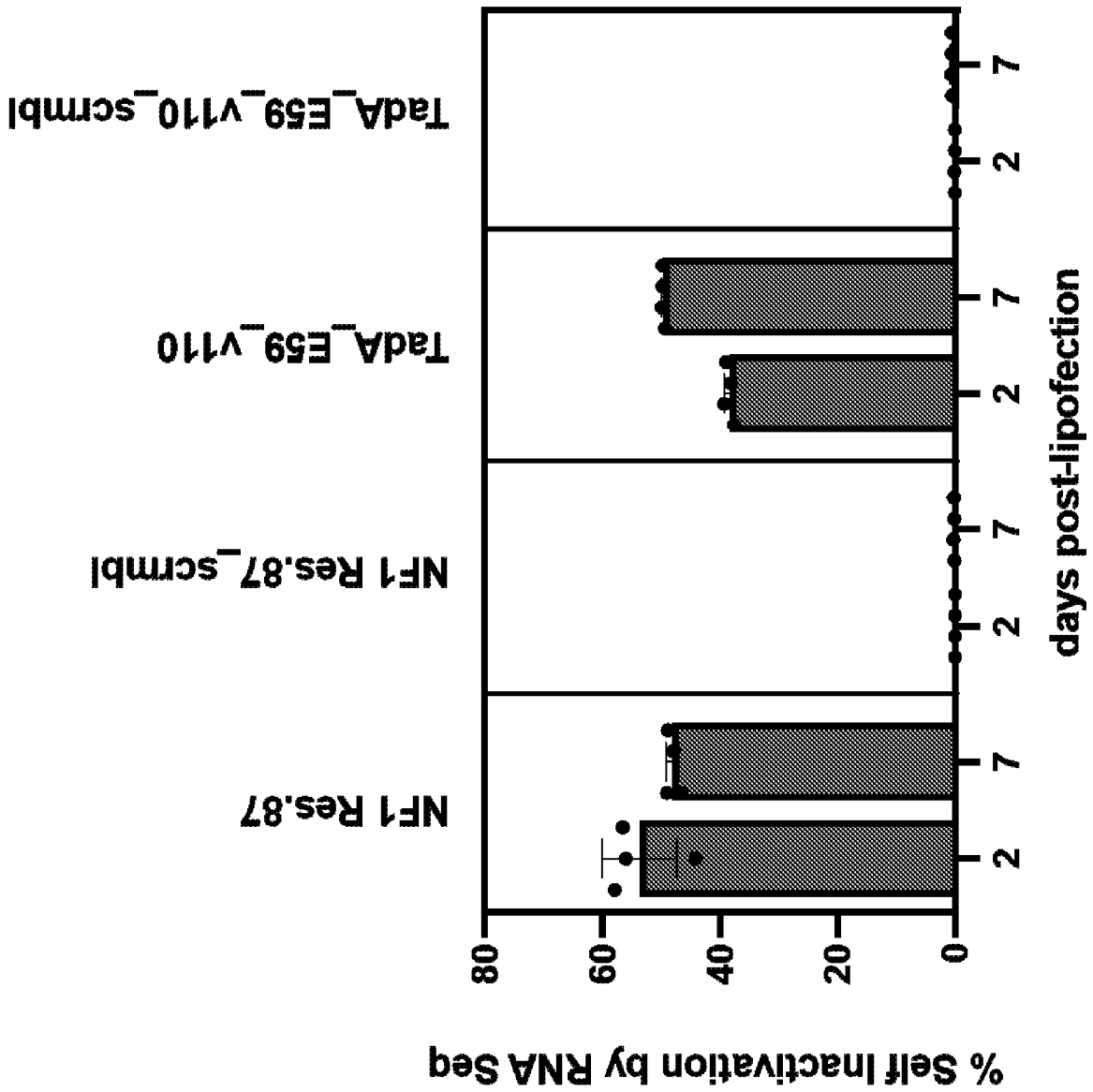

% Self Inactivation by RNA Seq
0
20
40
60
80
2
7
2
7
2
7
2
7
days post-lipofection
NF1 Res.87
NF1 Res.87_scrmbl
TadA_E59_V110
TadA_E59_V110_scrmbl

FIG. 7B

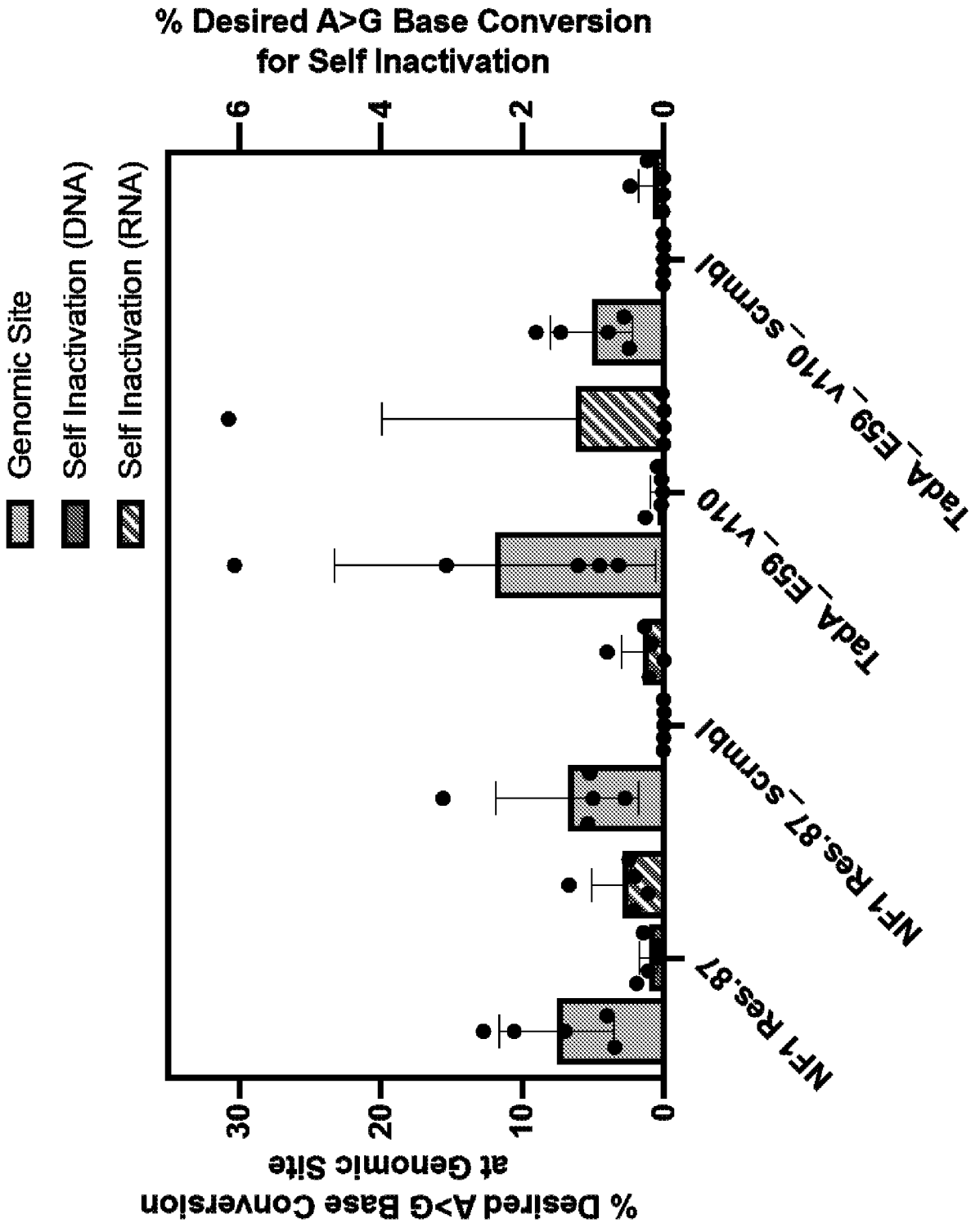

% Desired A>G Base Conversion
for Self Inactivation
Genomic Site
Self Inactivation (DNA)
Self Inactivation (RNA)
% Desired A>G Base Conversion
at Genomic Site
0
10
20
30
0
2
4
6
NF1 Res.87
NF1 Res.87_scrmbl
TadA_E59_V110
TadA_E59_V110_scrmbl

FIG. 8A

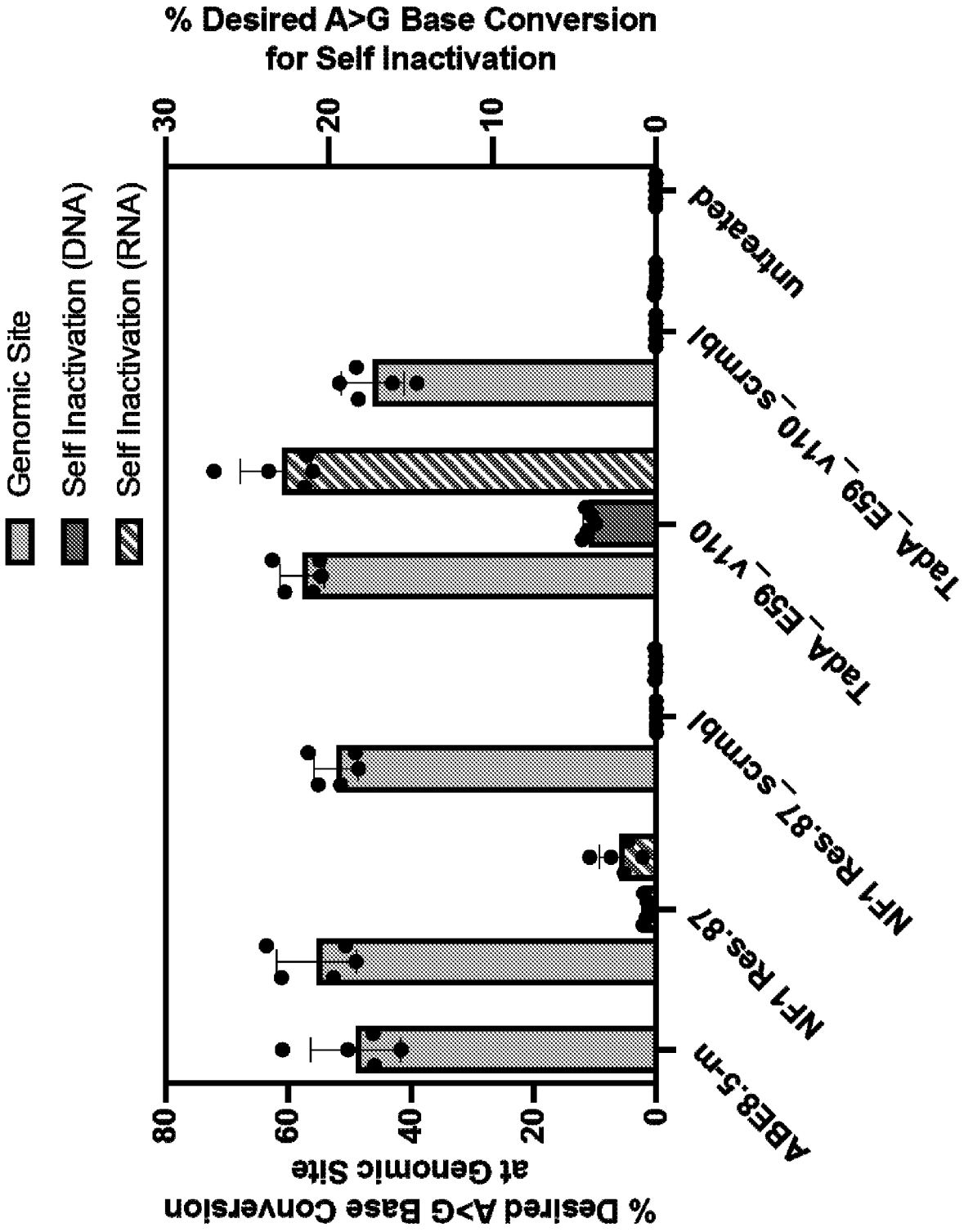

% Desired A>G Base Conversion
for Self Inactivation
Genomic Site
Self Inactivation (DNA)
Self Inactivation (RNA)
30
20
10
0
untreated
TadA_E59_V110_Scrmbl
TadA_E59_V110
NF1 Res.87_Scrmbl
NF1 Res.87
ABE8.5-m
80
60
40
20
0
% Desired A>G Base Conversion
at Genomic Site

FIG. 8B