



(12) **DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION**

(13) **A1**

(86) Date de dépôt PCT/PCT Filing Date: 2020/09/09
(87) Date publication PCT/PCT Publication Date: 2021/03/18
(85) Entrée phase nationale/National Entry: 2022/03/07
(86) N° demande PCT/PCT Application No.: US 2020/049975
(87) N° publication PCT/PCT Publication No.: 2021/050571
(30) Priorités/Priorities: 2019/09/09 (US62/897,777);
2020/02/13 (US PCT/US2020/018195)

(51) Cl.Int./Int.Cl. *A61K 31/7088* (2006.01),
C12N 15/113 (2010.01), *C12N 15/62* (2006.01)
(71) Demandeur/Applicant:
BEAM THERAPEUTICS INC., US
(72) Inventeurs/Inventors:
GAUDELLI, NICOLE, US;
PACKER, MICHAEL, US
(74) Agent: NORTON ROSE FULBRIGHT CANADA
LLP/S.E.N.C.R.L., S.R.L.

(54) Titre : NOUVEAUX EDITEURS DE NUCLEOBASES ET LEURS PROCEDES D'UTILISATION
(54) Title: NOVEL NUCLEOBASE EDITORS AND METHODS OF USING SAME

(57) Abrégé/Abstract:

The invention features novel programmable nucleobase editors comprising adenosine deaminase domains and methods of using the same for polynucleotide editing. In some embodiments, programmable nucleobase editors edit a pathogenic mutation associated with a genetic disease.



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

(19) World Intellectual Property
Organization
International Bureau



(10) International Publication Number
WO 2021/050571 A1

(43) International Publication Date
18 March 2021 (18.03.2021)

(51) International Patent Classification:

A61K 31/7088 (2006.01) C12N 15/62 (2006.01)
C12N 15/113 (2010.01)

(21) International Application Number:

PCT/US2020/049975

(22) International Filing Date:

09 September 2020 (09.09.2020)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/897,777 09 September 2019 (09.09.2019) US
PCT/US2020/018195
13 February 2020 (13.02.2020) US

(71) Applicant: **BEAM THERAPEUTICS INC.** [US/US]; 26
Landsdowne Street, 2nd Floor, Cambridge, Massachusetts
02139 (US).

(72) Inventors: **GAUDELLI, Nicole**; C/o Beam Therapeutics
Inc., 26 Landsdowne Street, 2nd Floor, Cambridge, Massa-
chusetts 02139 (US). **PACKER, Michael**; C/o Beam Ther-
apeutics Inc., 26 Landsdowne Street, 2nd Floor, Cambridge,
Massachusetts 02139 (US).

(74) Agent: **HUNTER-ENSOR, PH.D., Melissa**; C/o Green-
berg Traurig, LLP, One International Place, Boston, Mass-
achusetts 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, IR, IS, IT, JO, JP, KE, KG, KH, KN,
KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD,
ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO,
NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW,
SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

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

Declarations under Rule 4.17:

- as to the applicant's entitlement to claim the priority of the
earlier application (Rule 4.17(iii))
- of inventorship (Rule 4.17(iv))

Published:

- with international search report (Art. 21(3))
- 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))

(54) Title: NOVEL NUCLEOBASE EDITORS AND METHODS OF USING SAME

(57) Abstract: The invention features novel programmable nucleobase editors comprising adenosine deaminase domains and methods of using the same for polynucleotide editing. In some embodiments, programmable nucleobase editors edit a pathogenic mutation associated with a genetic disease.

WO 2021/050571 A1

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2
CONTENANT LES PAGES 1 À 270

NOTE : Pour les tomes additionnels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 2
CONTAINING PAGES 1 TO 270

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

NOVEL NUCLEOBASE EDITORS AND METHODS OF USING SAME

CROSS REFERENCE TO RELATED APPLICATIONS

5 This application is an International PCT Application which claims priority to and benefit of U.S. Provisional Application No. 62/897,777, filed September 9, 2019; and which claims priority to International PCT Application No. PCT/US2020/018195, filed February 13, 2020, the contents of all of which are incorporated by reference herein in their entireties.

BACKGROUND OF THE INVENTION

10 Targeted editing of nucleic acid sequences, for example, the targeted cleavage or the targeted introduction of a specific modification into genomic DNA is a highly promising approach for the study of gene function and also has the potential to provide new therapies for human genetic diseases. Currently available base editors include cytidine base editors (e.g., BE4) that convert target C•G base pairs to T•A and adenine base editors (e.g.,
15 ABE7.10) that convert A•T to G•C. There is a need in the art for improved base editors capable of inducing modifications within a target sequence with greater specificity and efficiency.

SUMMARY OF THE INVENTION

As described below, the present invention features novel programmable nucleobase
20 editors comprising adenosine deaminase domains (e.g., TadA*9 or ABE9), and methods of using the same for polynucleotide editing. In some embodiments, ABE9 of the invention edits a polynucleotide, e.g., a polynucleotide comprising a pathogenic mutation associated with a genetic disease.

In an aspect, an adenosine deaminase comprising an alteration at an amino acid
25 position selected from the group consisting of 21, 23, 25, 38, 51, 54, 70, 71, 72, 73, 94, 124, 133, 139, 146, and 158 of SEQ ID NO: 1, or a corresponding alteration in another adenosine deaminase:

	10	20	30	40	50
	MSEVEFSHEY	WMRHALLAK	<u>R</u> <u>A</u> <u>R</u> <u>D</u> <u>E</u> REVPV	GAVLVLN <u>N</u> RV	IGEGWNRAIG
30	60	70	80	90	100
	<u>L</u> <u>H</u> <u>D</u> <u>P</u> TAHAEI	MALRQGGLV <u>M</u>	<u>Q</u> <u>N</u> <u>Y</u> RLIDATL	YVTFEPCVMC	AGAM <u>I</u> HSRIG
	110	120	130	140	150
	RVVFGVRNAK	TGAAGSLMDV	LHY <u>P</u> GMNHRV	EI <u>T</u> EGIL <u>A</u> <u>D</u> E	CAALL <u>C</u> YFFR
	160				

MPRQVFNAQK KAQSSTD (SEQ ID NO: 1) is provided. In an embodiment, the adenosine deaminase comprises an alteration selected from the group consisting of R21N, R23H, E25F, N38G, L51W, P54C, M70V, Q71M, N72K, Y73S, M94V, P124W, T133K, D139L, D139M, C146R, and A158K of SEQ ID NO: 1, or a corresponding alteration in another adenosine deaminase. In an embodiment, the adenosine deaminase further comprises a V82T alteration of SEQ ID NO: 1, or a corresponding alteration in another adenosine deaminase. In an embodiment, the adenosine deaminase comprises alterations at two or more amino acid positions selected from the group consisting of 21, 23, 25, 38, 51, 54, 70, 71, 72, 73, 94, 124, 133, 139, 146, and 158 of SEQ ID NO: 1, or a corresponding alteration in another adenosine deaminase. In an embodiment, the adenosine deaminase of this aspect and embodiments thereof comprises two or more of the alterations. In an embodiment, the adenosine deaminase of this aspect and embodiments thereof comprises three or more of said alterations. In an embodiment, the adenosine deaminase of this aspect and embodiments thereof further comprises one or more of the following alterations: Y147T, Y147R, Q154S, Y123H, and Q154R. In an embodiment, the adenosine deaminase of this aspect and embodiments thereof comprises any one of the following groups 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;
 5 E25F + V82S + Y123H + D139M + Y147R + Q154R;
 M70V + V82S + M94V + Y123H + Y147R + Q154R;
 Q71M + V82S + Y123H + Y147R + Q154R;
 E25F + I76Y + V82S + Y123H + Y147R + Q154R;
 I76Y + V82T + Y123H + Y147R + Q154R;
 10 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;
 15 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;
 20 P54C + I76Y + V82S + Y123H + Y147R + Q154R;
 R21N + I76Y + V82S + Y123H + Y147R + Q154R;
 I76Y + V82S + Y123H + D139M + Y147R + Q154R;
 Y73S + I76Y + V82S + Y123H + Y147R + Q154R;
 V82S + Q154R;
 25 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;
 30 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; or
 M70V + Q71M + N72K + V82S + Y123H + Y147R + Q154R. In an embodiment, the
 adenosine deaminase variant comprises any alteration or group of alterations as described in
 Table 14 or 18. In an embodiment, the adenosine deaminase of this aspect and embodiments
 5 thereof comprises a deletion of the C terminus beginning at a residue selected from the group
 consisting of 149, 150, 151, 152, 153, 154, 155, 156, and 157. In an embodiment, the
 adenosine deaminase of this aspect and embodiments thereof further comprises an alteration
 selected from the group consisting of Y147T, Y147R, Q154S, Y123H, V82S, T166R, and
 Q154R. In an embodiment, the adenosine deaminase of this aspect and embodiments thereof
 10 is an adenosine deaminase variant described in Table 14, Table 18, or FIGS. 3A-3C.

In another aspect, a fusion protein is provided, in which the fusion protein comprises a
 polynucleotide programmable DNA binding domain and at least one base editor domain that
 is an adenosine deaminase variant comprising an alteration at an amino acid position selected
 from the group consisting of 21, 23, 25, 38, 51, 54, 70, 71, 72, 73, 94, 124, 133, 139, 146, and
 15 158 of the below SEQ ID NO: 1, or a corresponding alteration in another adenosine
 deaminase:

	10	20	30	40	50
	MSEVEFSHEY	WMRHALLTAK	<u>R</u> <u>A</u> <u>R</u> <u>D</u> <u>E</u> REVPV	GAVLVLN <u>N</u> RV	IGEGWNRAIG
	60	70	80	90	100
20	<u>L</u> <u>H</u> <u>D</u> <u>P</u> TAHAEI	MALRQGGLV <u>M</u>	<u>Q</u> <u>N</u> <u>Y</u> RLIDATL	YVTFEPCVMC	AGAM <u>I</u> HSRIG
	110	120	130	140	150
	RVVFGVRNAK	TGAAGSLMDV	LHY <u>P</u> GMNHRV	EI <u>T</u> EGILADE	CAALL <u>C</u> YFFR
	160				

MPRQVFNAQK KAQSSTD (SEQ ID NO: 1). In an embodiment, the adenosine deaminase
 25 variant comprises an alteration selected from the group consisting of R21N, R23H, E25F,
 N38G, L51W, P54C, M70V, Q71M, N72K, Y73S, M94V, P124W, T133K, D139L, D139M,
 C146R, and A158K of SEQ ID NO: 1, or a corresponding alteration in another adenosine
 deaminase.

In another aspect, a fusion protein is provided, in which the fusion protein comprises a
 30 polynucleotide programmable DNA binding domain and at least one base editor domain that
 is an adenosine deaminase variant comprising an alteration selected from the group consisting
 of R21N, R23H, E25F, N38G, L51W, P54C, M70V, Q71M, N72K, Y73S, M94V, P124W,
 T133K, D139L, D139M, C146R, and A158K of SEQ ID NO: 1, or a corresponding alteration
 in another adenosine deaminase.

In an embodiment of any of fusion protein of any of the above-delineated aspects and embodiments thereof, the adenosine deaminase variant further comprises a V82T alteration of SEQ ID NO: 1, or a corresponding alteration in another adenosine deaminase.

In another aspect, a fusion protein is provided, in which the fusion protein comprises a polynucleotide programmable DNA binding domain and at least one base editor domain that is an adenosine deaminase variant comprising an alteration V82T and one or more alterations selected from the group consisting of R21N, R23H, E25F, N38G, L51W, P54C, M70V, Q71M, N72K, Y73S, M94V, P124W, T133K, D139L, D139M, C146R, and A158K of SEQ ID NO: 1, or a corresponding alteration in another adenosine deaminase.

In an embodiment of the fusion proteins of any of the above-delineated aspects and embodiments thereof, the adenosine deaminase variant comprises alterations at two or more amino acid positions selected from the group consisting of 21, 23, 25, 38, 51, 54, 70, 71, 72, 73, 94, 124, 133, 139, 146, and 158 of SEQ ID NO: 1, or a corresponding alteration in another adenosine deaminase. In an embodiment, the adenosine deaminase variant comprises two or more of the alterations. In an embodiment, the adenosine deaminase variant comprises three or more of the alterations. In an embodiment, the adenosine deaminase variant further comprises one or more of the following alterations: Y147T, Y147R, Q154S, Y123H, and Q154R. In an embodiment, the adenosine deaminase variant comprises a deletion of the C terminus beginning at a residue selected from the group consisting of 149, 150, 151, 152, 153, 154, 155, 156, and 157.

In an embodiment of the above-delineated fusion proteins and embodiments thereof, the base editor domain comprises an adenosine deaminase variant monomer, wherein the adenosine deaminase monomer comprises one or more alterations selected from the group consisting of R21N, R23H, E25F, N38G, L51W, P54C, M70V, Q71M, N72K, Y73S, V82T, M94V, P124W, T133K, D139L, D139M, C146R, and A158K of SEQ ID NO: 1. In an embodiment, the base editor domain comprises an adenosine deaminase heterodimer comprising a wild-type adenosine deaminase domain and an adenosine deaminase variant. In an embodiment, the adenosine deaminase variant further comprises an alteration selected from the group consisting of Y147T, Y147R, Q154S, Y123H, V82S, T166R, and Q154R. In an embodiment, the base editor domain comprises an adenosine deaminase heterodimer comprising a TadA*7.10 domain and adenosine deaminase variant domain. In an embodiment, the adenosine deaminase variant comprises two or more alterations.

In another embodiment of the fusion proteins of any of the above-delineated aspects and embodiments thereof, the adenosine deaminase variant is an ABE9 (TadA*9 deaminase variant) described in Table 14, Table 18, or FIGS. 3A-3C.

In another embodiment of the fusion proteins of any of the above-delineated aspects and embodiments thereof, the adenosine deaminase variant is a truncated ABE8 or ABE9 that is missing 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 C-terminal amino acid residues relative to the full length ABE9.

In another embodiment of the fusion proteins of any of the above-delineated aspects and embodiments thereof, the polynucleotide programmable DNA binding domain is a Cas9, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ domain.

In another aspect, a fusion protein is provided, in which the fusion protein comprises a polynucleotide programmable DNA binding domain comprising the following sequence:

EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
VRKVLSPMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPPKKYGGFM
QPTVAYSVLVVAKEKGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKE
VKKDLIIKLPKYSLEFLENGRKRMLASAKFLQKGNELALPSKYVNFYLASHYE
KLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPRAFKYFDTTARKEYRSTKEVLDTLIHQSITE
LYETRIDLSQLGGDGGSGGSGGSGGSGGSGGSGGMDKKYSIGLAIGTNSVGWAVI
TDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR
KNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKL
FIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFG
NLIAISLGLTPNFKSNFDLAEDAKQLSKDQYDDDLNLLAQIGDQYADLFLAA
KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLEPKYK
EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPPFLKDNREKIEKILTFRIPIYYVGPLARG
NSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMNTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK
EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKHKDKDFLDNEENEDILEDIVL
TLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQ
SGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA

GSPA**IKKGILQTVKVVDELVKVMGRHKPENIV****EMARENQTTQKGQKNSRERM**
KRIE**EGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSD**
YD**VDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLN**
AKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTK
 5 **YDENDKLIREVKVITLKS****KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTA**
LIK**KYPKLESEFVYGDYKVYDVRK****MI****AKSE****QEGADKRTADGSEFESPKKKRKV***

wherein the bold sequence indicates sequence derived from Cas9, the italics sequence denotes a linker sequence, and the underlined sequence denotes a bipartite nuclear localization sequence, and at least one base editor domain comprising an adenosine

- 10 deaminase variant comprising an alteration at an amino acid position selected from the group consisting of 21, 23, 25, 38, 51, 54, 70, 71, 72, 73, 94, 124, 133, 138, 139, 146, and 158 of SEQ ID NO: 1. In an embodiment, the adenosine deaminase variant comprises an alteration selected from the group consisting of R21N, R23H, E25F, N38G, L51W, P54C, M70V, Q71M, N72K, Y73S, M94V, P124W, T133K, D138M, D139L, D139M, C146R, and A158K
- 15 of SEQ ID NO: 1. In another embodiment, the adenosine deaminase variant comprises an alteration V82T of SEQ ID NO: 1. In an embodiment, the adenosine deaminase variant comprises two or more of said alterations. In an embodiment, the adenosine deaminase variant comprises three or more of said alterations. In an embodiment, the adenosine deaminase variant further comprises an alteration selected from the group consisting of
- 20 Y147T, Y147R, Q154S, Y123H, V82S, T166R, and Q154R. In an embodiment, the adenosine deaminase variant comprises two or more of the following alterations: Y147T, Y147R, Q154S, Y123H, and Q154R.

In an embodiment of any of the above-delineated fusion proteins and embodiments thereof, the adenosine deaminase variant comprises any one of the following groups of

25 alterations:

E25F + V82S + Y123H;

T133K + Y147R + Q154R;

E25F + V82S + Y123H + Y147R + Q154R;

L51W + V82S + Y123H + C146R + Y147R + Q154R;

- 30 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;
 5 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;
 10 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;
 15 M70V + V82S + M94V + Y123H + Y147R + Q154R;
 Q71M + V82S + Y123H + Y147R + Q154R;
 E25F + I76Y + V82S + Y123H + Y147R + Q154R;
 I76Y + V82T + Y123H + Y147R + Q154R;
 N38G + I76Y + V82S + Y123H + Y147R + Q154R;
 20 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;
 25 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;
 30 R21N + I76Y + V82S + Y123H + Y147R + Q154R;
 I76Y + V82S + Y123H + D139M + Y147R + Q154R;
 Y73S + I76Y + V82S + Y123H + Y147R + Q154R;
 V82S + Q154R;

N72K + V82S + Y123H + Y147R + Q154R;

Q71M + V82S + Y123H + Y147R + Q154R;

V82S + Y123H + T133K + Y147R + Q154R;

V82S + Y123H + T133K + Y147R + Q154R + A158K;

5 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;

10 V82S + Y123H + T133K + Y147R + Q154R + A158K;

M70V + Q71M + N72K + V82S + Y123H + Y147R + Q154R.

In an embodiment, the adenosine deaminase variant comprises any other alteration or group of alterations as described in Table 14 or 18, or in FIGS. 3A-3C.

In an embodiment of the fusion proteins of any of the above-delineated aspects and
15 embodiments thereof, 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 an embodiment of the fusion proteins of any of the above-delineated aspects and
embodiments thereof, the polynucleotide programmable DNA binding domain comprises a
20 modified SaCas9 having an altered protospacer-adjacent motif (PAM) specificity. In an
embodiment, the modified SaCas9 comprises amino acid substitutions E782K, N968K, and
R1015H, or corresponding amino acid substitutions thereof.

In an embodiment of the fusion proteins of any of the above-delineated aspects and
embodiments thereof, the polynucleotide programmable DNA binding domain comprises a
25 variant of SpCas9 having an altered protospacer-adjacent motif (PAM) specificity. In an
embodiment, the altered PAM has specificity for the nucleic acid sequence 5'-NGA-3', 5'-
NGC-3', 5'-NGG-3', 5'-NGT-3', or 5'-NGN-3'. In an embodiment, the variant SpCas9
comprises amino acid substitutions selected from: D1135M, S1136Q, G1218K, E1219F,
A1322R, D1332A, R1335E, and T1337R, or corresponding amino acid substitutions thereof;
30 I322V, S409I, E427G, R654L, R753G (MQKFRAER) or corresponding amino acid
substitutions thereof; I322V, S409I, E427G, R654L, R753G, R1114G, or corresponding
amino acid substitutions thereof; or amino acid substitutions as set forth in FIGS. 3A-3C.

In an embodiment of the fusion proteins of any of the above-delineated aspects and embodiments thereof, the polynucleotide programmable DNA binding domain is a nuclease inactive or nickase variant. In an embodiment, the nickase variant comprises an amino acid substitution D10A or a corresponding amino acid substitution thereof.

5 In an embodiment of the fusion proteins of any of the above-delineated aspects and embodiments thereof, the adenosine deaminase domain is capable of deaminating adenine in deoxyribonucleic acid (DNA).

In an embodiment of the fusion proteins of any of the above-delineated aspects and embodiments thereof, the adenosine deaminase is a modified adenosine deaminase that does
10 not occur in nature.

In an embodiment of the adenosine deaminase of the above-delineated aspect and embodiments thereof, the adenosine deaminase is a TadA deaminase. In an embodiment of the fusion proteins of any of the above-delineated aspects and embodiments thereof, the adenosine deaminase is a TadA deaminase. In an embodiment, the TadA deaminase is a
15 TadA*7.10 variant.

In an embodiment of the fusion proteins of any of the above-delineated aspects and embodiments thereof, the fusion protein comprises a linker between the polynucleotide programmable DNA binding domain and the adenosine deaminase domain. In an embodiment, the linker comprises the amino acid sequence:
20 SGGSSGGSSGSETPGTSESATPES.

In an embodiment of the fusion proteins of any of the above-delineated aspects and embodiments thereof, the fusion proteins comprises one or more nuclear localization signals. In an embodiment, the nuclear localization signal is a bipartite nuclear localization signal.

In an embodiment of the fusion proteins of any of the above-delineated aspects and
25 embodiments thereof, the Cas9 is a StCas9.

In an embodiment of the fusion proteins of any of the above-delineated aspects and embodiments thereof, the Cas9 is a SaCas9 or an SpCas9.

In an embodiment of the fusion proteins of any of the above-delineated aspects and embodiments thereof, the Cas9 is a modified SaCas9 or a modified SpCas9. In an
30 embodiment, the modified SaCas9 comprises amino acid substitutions E782K, N968K, and R1015H, or corresponding amino acid substitutions thereof. In an embodiment, the modified SaCas9 comprises the amino acid sequence:

KRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRR
 RRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRR
 GVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKT
 SDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEW
 5 YEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYEYEFQIEN
 VFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIENA
 ELLDQIAKILTIYQSSEDIQEELTNLNSLTQEEIEQISNLKGYTGTHNLSLKAINLILDE
 LWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIK
 KYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIRTTGKENAKYLIEKIK
 10 LHDMQEGKCLYSLEAIPLEDLLNPNFNYEVDHIIPRSVSFDNSFNKVLVKQEENSKK
 GNRTPFQYLSSSDSKISYETFKKHILNLAQKGRISKTKKEYLLEERDINRFSVQKDFI
 NRNLVDTRYATRGLMNLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKG
 YKHHAEADALIANADFIFKEWKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIF
 ITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNNLNGLYDK
 15 DNDKLKKLINKSPEKLLMYHHPQTYQKLKLIMEQYGDEKNPLYKYEEETGNYLTK
 YSKKDNGPVIKKIKYYGNKLNALDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVY
 KFVTVKNLDVIKKENYEEVNSKCYEEAKKLKKISNQAEFIASFYKNDLIKINGELYR
 VIGVNNDLLNRIEVMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLY
 EVKSKKHPQIIKKG.

20 In another aspect, a polynucleotide encoding the fusion protein of any one of the
 above-delineated aspects and embodiments thereof is provided.

In another aspect, a cell is provided, in which the cell is produced by introducing into
 the cell, or a progenitor thereof: a polynucleotide encoding the fusion protein of any one of
 the above-delineated aspects and embodiments thereof, and one or more guide
 25 polynucleotides that target the base editor to effect an A•T to G•C alteration of a SNP
 associated with a genetic disease. In an embodiment, the cell is a human cell. In an
 embodiment, the cell is *in vitro* or *in vivo*. In an embodiment, the genetic disease is alpha-1
 antitrypsin deficiency (A1AD). In an embodiment, the fusion protein and the one or more
 guide polynucleotides forms a complex in the cell.

30 In another aspect, an isolated cell or population of cells propagated or expanded from
 the cell of the above-delineated aspect and embodiments thereof is provided.

In an aspect, a method of treating a genetic disease in a subject in need thereof is
 provided, in which the method comprises administering to the subject the cell, isolated cell,

or population of cells of any one of the above-delineated aspects and embodiments thereof. In an embodiment of the method, the cell, isolated cell, or population of cells is autologous, allogeneic, or xenogeneic to the subject.

In an aspect, a base editor system is provided, in which the base editor system comprises a polynucleotide programmable DNA binding domain and at least one base editor domain that is an adenosine deaminase variant comprising an alteration at an amino acid position selected from the group consisting of 21, 23, 25, 38, 51, 54, 70, 71, 72, 73, 82, 94, 124, 133, 139, 146, and 158 of the following SEQ ID NO: 1, a corresponding alteration in another adenosine deaminase:

```

10      10      20      30      40      50
MSEVEFSHEY WMRHALTLAK RARDEREVVPV GAVLVLNNRV IGEGWNRRAIG
      60      70      80      90      100
LHDPTAHA EI MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG
      110     120     130     140     150
15 RVVFGVRNAK TGAAGSLMDV LHYPG MNHRV EITEGILADE CAALLCYFFR
      160

```

MPRQVFNAAQK KAQSSTD (SEQ ID NO: 1). In an embodiment of the base editor system, the adenosine deaminase variant comprises an alteration selected from the group consisting of R21N, R23H, E25F, N38G, L51W, P54C, M70V, Q71M, N72K, Y73S, V82T, M94V, P124W, T133K, D139L, D139M, C146R, and A158K of SEQ ID NO: 1, or a corresponding alteration in another adenosine deaminase. In an embodiment, the base editor system further comprises one or more guide polynucleotides that target the base editor domain to effect an A•T to G•C alteration of a SNP associated with a genetic disease. In an embodiment, of the base editor system, the adenosine deaminase variant is capable of deaminating adenine in deoxyribonucleic acid (DNA). In an embodiment of the base editor system, the guide polynucleotide comprises ribonucleic acid (RNA), or deoxyribonucleic acid (DNA). In an embodiment of the base editor system, the guide polynucleotide comprises a CRISPR RNA (crRNA) sequence, a trans-activating CRISPR RNA (tracrRNA) sequence, or a combination thereof. In an embodiment, the base editor system further comprises a second guide polynucleotide. In an embodiment, the second guide polynucleotide comprises ribonucleic acid (RNA), or deoxyribonucleic acid (DNA). In an embodiment, the second guide polynucleotide comprises a CRISPR RNA (crRNA) sequence, a trans-activating CRISPR RNA (tracrRNA) sequence, or a combination thereof. In an embodiment of the above-delineated base editor system and embodiments thereof, the polynucleotide-programmable DNA-binding domain comprises a Cas9, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ domain. In an

embodiment, the polynucleotide-programmable DNA-binding domain is nuclease dead. In an embodiment, the polynucleotide-programmable DNA-binding domain is a nickase. In an embodiment, the polynucleotide-programmable DNA-binding domain comprises a Cas9 domain. In an embodiment, the Cas9 domain comprises a nuclease dead Cas9 (dCas9), a Cas9 nickase (nCas9), or a nuclease active Cas9. In an embodiment, the Cas9 domain comprises a Cas9 nickase. In an embodiment, the polynucleotide-programmable DNA-binding domain is an engineered or a modified polynucleotide-programmable DNA-binding domain. In an embodiment of the above-delineated base editor system and embodiments thereof, the genetic disease is alpha-1 antitrypsin deficiency (A1AD).

In another aspect, a method for correcting a single nucleotide polymorphism (SNP) in a polynucleotide is provided, in which the method comprises: contacting a target nucleotide sequence, at least a portion of which is located in the polynucleotide or its reverse complement, with a fusion protein of any one of the above-delineated aspects and embodiments thereof, or the base editor system of any one of the above-delineated aspects and embodiments thereof; and editing the SNP by deaminating the SNP or its complement nucleobase upon targeting of the base editor to the target nucleotide sequence, wherein deaminating the SNP or its complement nucleobase corrects the SNP. In an embodiment, the SNP is associated with alpha-1 antitrypsin deficiency (A1AD). In an embodiment, the SNP is in the *SERPINA1* gene and the correction comprises an E342K (PiZ allele) alteration.

In an aspect, a method for editing a polynucleotide is provided, in which the method comprises contacting a target nucleotide sequence with the fusion protein of any one of the above-delineated aspects and embodiments thereof, or the base editor system of any one of the above-delineated aspects and embodiments thereof, thereby editing the polynucleotide. In an embodiment of the method, the editing results in less than 20% indel formation, less than 15% indel formation, less than 10% indel formation; less than 5% indel formation; less than 4% indel formation; less than 3% indel formation; less than 2% indel formation; less than 1% indel formation; less than 0.5% indel formation; or less than 0.1% indel formation. In an embodiment of the method, the editing does not result in translocations.

In another aspect is provided a base editor comprising an ABE9 (TadA*9 deaminase variant) comprising a TadA*7.10 adenosine deaminase variant domain and a Cas9 endonuclease domain selected from the following:

monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+A109S of SEQ ID NO: 1, and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER);

monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+T111R of SEQ ID NO: 1, and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER);

monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+D119N of SEQ ID NO: 1, and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER);

5 monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+H122N of SEQ ID NO: 1, and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER);

monoTadA*7.10 having mutations I76Y+V82T+Y147d+Q154S of SEQ ID NO: 1, and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER);

10 monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+F149Y of SEQ ID NO: 1, and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER);

monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+T166I of SEQ ID NO: 1, and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER); and

15 monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+D167N of SEQ ID NO: 1, and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER).

mono TadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+L36H+N157K of SEQ ID NO: 1, and spCas9 having mutations I322V, S409I, E427G, R654L, R753G, R1114G (MQKFRAER);

20 mono TadA*7.10 having mutations I76Y+V82T+Y147D+Q154S+F149Y+D167N+L36H+N157K of SEQ ID NO: 1, and SpCas9 having mutations I322V, S409I, E427G, R654L, R753G, R1114G (MQKFRAER);

monoTadA*7.10 having mutations I76Y+V82T+Y147D+Q154S+F149Y+D167N+L36H+N157K+V106W of SEQ ID NO: 1, and SpCas9 having mutations I322V, S409I, E427G, R654L, R753G, R1114G (MQKFRAER);

mono TadA*7.10 having mutations A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N of SEQ ID NO: 1, and SpCas9 having mutations I322V, S409I, E427G, R654L, R753G, R1114G, MQKFRAER; and

30 mono TadA*7.10 having mutations A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N + V106W of SEQ ID NO: 1, and SpCas9 having mutations I322V, S409I, E427G, R654L, R753G, R1114G (MQKFRAER); and one or more guide polynucleotides that target the adenosine deaminase variant domain to effect an A•T to G•C

alteration of a SNP associated with a genetic disease. In an embodiment of the base editor, the SNP is associated with alpha-1 antitrypsin deficiency (A1AD).

In another aspect, a vector is provided in which the vector comprises one or more polynucleotides encoding an ABE9 base editor comprising a TadA adenosine deaminase
5 domain and an SpCas9 endonuclease domain selected from

monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+A109S and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER);

monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+T111R and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER);

10 monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+D119N and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER);

monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+H122N and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER);

15 monoTadA*7.10 having mutations I76Y+V82T+Y147d+Q154S and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER);

monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+F149Y and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER);

monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+T166I and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER); and

20 monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+D167N and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER).

mono TadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+L36H+N157K and spCas9 having mutations I322V, S409I, E427G, R654L, R753G, R1114G (MQKFRAER);

25 mono TadA*7.10 having mutations I76Y+V82T+Y147D+Q154S+F149Y+D167N+L36H+N157K and SpCas9 having mutations I322V, S409I, E427G, R654L, R753G, R1114G (MQKFRAER);

monoTadA*7.10 having mutations I76Y+V82T+Y147D+Q154S+F149Y+D167N+L36H+N157K+V106W and SpCas9 having mutations I322V, S409I, E427G, R654L, R753G, R1114G, (MQKFRAER)

30 mono TadA*7.10 having mutations A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N and SpCas9 having mutations I322V, S409I, E427G, R654L, R753G, R1114G (MQKFRAER); and

mono TadA*7.10 having mutations A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N + V106W and SpCas9 having mutations I322V, S409I, E427G, R654L, R753G, R1114G (MQKFRAER). In an embodiment, the vector is a plasmid, viral, or mRNA vector.

5 In another aspect, a composition is provided, in which the composition comprises the fusion protein of any one of the above-delineated aspects and embodiments thereof or the base editor system of any one of the above-delineated aspects and embodiments thereof. In an embodiment, the composition further comprises a pharmaceutically acceptable excipient, diluent, or carrier.

10 In another aspect, a composition comprising the fusion protein of any one of the above-delineated aspects and embodiments thereof bound to a guide RNA is provided, wherein the guide RNA comprises a nucleic acid sequence that is complementary to an *SERPINA1* gene associated with alpha-1 antitrypsin deficiency (A1AD).

15 In another aspect, a composition comprising the base editor system of any one of the above-delineated aspects and embodiments thereof bound to a guide RNA is provided, wherein the guide RNA comprises a nucleic acid sequence that is complementary to an *SERPINA1* gene associated with alpha-1 antitrypsin deficiency (A1AD).

In an embodiment of the compositions of any one of the above-delineated aspects and embodiments thereof, the adenosine deaminase variant is capable of deaminating adenine in
20 deoxyribonucleic acid (DNA).

In an embodiment of the compositions of any one of the above-delineated aspects and embodiments thereof, the fusion protein or base editor system

(i) comprises a Cas9 nickase;

(ii) comprises a nuclease inactive Cas9;

25 (iii) comprises an SpCas9 variant comprising a combination of amino acid substitutions shown in FIGS. 3A-3C; or

(iv) comprises an SpCas9 variant comprising a combination of amino acid sequence substitutions selected from I322V, S409I, E427G, R654L, R753G (MQKFRAER); or I322V, S409I, E427G, R654L, R753G, R1114G, (MQKFRAER).

30 In an embodiment of the compositions of any one of the above-delineated aspects and embodiments thereof, the composition further comprises a pharmaceutically acceptable excipient, diluent, or carrier, i.e., a pharmaceutical composition.

In an aspect, a pharmaceutical composition for the treatment of a disease or disorder comprising the composition further comprising a pharmaceutically acceptable excipient, diluent, or carrier is provided. In an embodiment of the pharmaceutical composition, the disease or disorder is alpha-1 antitrypsin deficiency (A1AD). In an embodiment of the pharmaceutical composition, the fusion protein or the base editor system is bound to a guide RNA, wherein the guide RNA comprises a nucleic acid sequence that is complementary to an *SERPINA1* gene associated with alpha-1 antitrypsin deficiency (A1AD). In an embodiment of the pharmaceutical composition, the gRNA and the base editor are formulated together or separately. In an embodiment of the above-delineated pharmaceutical composition and

embodiments thereof, the gRNA comprises a nucleic acid sequence, from 5' to 3', or a 1, 2, 3, 4, or 5 nucleotide 5' truncation fragment thereof, selected from one or more of

5'-ACCAUCGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC
AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU
CGGUGCUUUU-3';

5'-CCAUCGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC
AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU
CGGUGCUUUU-3';

5'-CAUCGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU
AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3';

5'-AUCGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU
AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3';

5'-UCGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU
AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3'; or

5'-CGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU
AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3'. In an

embodiment of the above-delineated pharmaceutical composition and embodiments thereof, the pharmaceutical composition further comprises a vector suitable for expression in a mammalian cell, wherein the vector comprises a polynucleotide encoding the base editor. In an embodiment of the pharmaceutical composition, the polynucleotide encoding the base editor is mRNA. In an embodiment of the pharmaceutical composition, the vector is a viral vector. In an embodiment of the pharmaceutical composition, the viral vector is a retroviral vector, adenoviral vector, lentiviral vector, herpesvirus vector, or adeno-associated viral vector (AAV). In an embodiment of the pharmaceutical composition of any one of the

above-delineated aspects and embodiments thereof, the pharmaceutical composition further comprises a ribonucleoparticle suitable for expression in a mammalian cell. In an embodiment of the pharmaceutical composition of any one of the above-delineated aspects and embodiments thereof, the pharmaceutical composition further comprises a lipid.

5 In another aspect, a method of treating alpha-1 antitrypsin deficiency (A1AD) is provided, in which the method comprises administering to a subject in need thereof the pharmaceutical composition of any one of the above-delineated aspects and embodiments thereof.

10 In another aspect, use of the pharmaceutical composition of any one of the above-delineated aspects and embodiments thereof in the treatment of alpha-1 antitrypsin deficiency (A1AD) in a subject is provided.

In an embodiment of the above-delineated method or use, the subject is a human.

In an embodiment of the fusion protein or base editor system of any one of the above-delineated aspects and embodiments thereof, the adenosine deaminase variant comprises any
15 one of the following groups of alterations:

E25F + V82S + Y123H;

T133K + Y147R + Q154R;

E25F + V82S + Y123H + Y147R + Q154R;

L51W + V82S + Y123H + C146R + Y147R + Q154R;

20 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;

25 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;

30 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;
5 M70V + V82S + M94V + Y123H + Y147R + Q154R;
Q71M + V82S + Y123H + Y147R + Q154R;
E25F + I76Y + V82S + Y123H + Y147R + Q154R;
I76Y + V82T + Y123H + Y147R + Q154R;
N38G + I76Y + V82S + Y123H + Y147R + Q154R;
10 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;
15 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;
20 R21N + I76Y + V82S + Y123H + Y147R + Q154R;
I76Y + V82S + Y123H + D139M + Y147R + Q154R;
Y73S + I76Y + V82S + Y123H + Y147R + Q154R;
V82S + Q154R;
N72K + V82S + Y123H + Y147R + Q154R;
25 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;
30 Q71M + V82S + Y123H + Y147R + Q154R;
M70V + V82S + M94V + Y123H + Y147R + Q154R;
V82S + Y123H + T133K + Y147R + Q154R;
V82S + Y123H + T133K + Y147R + Q154R + A158K;

M70V + Q71M + N72K + V82S + Y123H + Y147R + Q154R.

In an embodiment, the adenosine deaminase variant, e.g., TadA*9 deaminase variant) comprises any alteration or group of alterations as described in Table 14 or 18.

As would be appreciated by the skilled practitioner in the art in connection with the adenosine deaminases of the above-delineated aspects and embodiments thereof, amino acid alterations in other adenosine deaminases, which correspond to the amino acid alterations set forth in SEQ ID NO: 1, may be readily determined by performing routine sequence alignments and assessing relatedness and/or identities of the amino acid sequence of SEQ ID NO: 1 and the sequences, or relevant portions thereof, of other adenosine deaminase(s), such as TadA deaminases and the like, as described *supra*. In an embodiment, the amino acid sequence of another adenosine deaminase comprises at least 85% sequence identity to SEQ ID NO:1. In an embodiment, the amino acid sequence of another adenosine deaminase comprises at least 90% sequence identity to SEQ ID NO:1. In an embodiment, the amino acid sequence of another adenosine deaminase comprises at least 95% sequence identity to SEQ ID NO:1. In an embodiment, the amino acid sequence of another adenosine deaminase comprises at least 98% sequence identity to SEQ ID NO:1. In an embodiment, the amino acid sequence of another adenosine deaminase comprises at least 99% sequence identity to SEQ ID NO:1.

In another aspect is provided the above-delineated adenosine deaminase, fusion protein, base editor, or base editor system and embodiments thereof, comprising the adenosine deaminase or adenosine deaminase variant, which is a TadA*7.10 variant comprising any one of the following amino acid alterations or groups of alterations: V82T; I76Y + V82T; or I76Y + V82T + Y147T + Q154S.

In another aspect is provided an adenosine deaminase variant which is a TadA*7.10 variant comprising any one of the following amino acid alterations or groups of alterations: V82T; I76Y + V82T; or I76Y + V82T + Y147T + Q154S.

In another aspect, a fusion protein is provided, in which the fusion protein comprises a polynucleotide programmable DNA binding domain and at least one base editor domain that is an TadA*7.10 adenosine deaminase variant comprising any one of the following amino acid alterations or groups of alterations: V82T; I76Y + V82T; or I76Y + V82T + Y147T + Q154S. In an embodiment of the fusion protein, the polynucleotide programmable DNA binding domain comprises a Cas9 endonuclease domain. In an embodiment of the fusion

protein, the Cas9 endonuclease domain comprises spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER).

In an embodiment of the above-delineated adenosine deaminase variant and embodiments thereof, or the above-delineated fusion protein and embodiments thereof, the
5 TadA*7.10 is monomeric.

In another aspect, a nucleobase editor is provided in which the nucleobase editor comprises a TadA*7.10 adenosine deaminase variant domain and a Cas9 endonuclease domain selected from the following:

monoTadA*7.10 having mutation V82T and spCas9 having mutations I322V, S409I,
10 E427G, R654L, R753G (MQKFRAER);

monoTadA*7.10 having mutations I76Y + V82T and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER); or

monoTadA*7.10 having mutations I76Y + V82T + Y147T + Q154S and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER).

15

Definitions

The following definitions supplement those in the art and are directed to the current application and are not to be imputed to any related or unrelated case, *e.g.*, to any commonly owned patent or application. Although any methods and materials similar or equivalent to
20 those described herein can be used in the practice for testing of the present disclosure, the preferred materials and methods are described herein. Accordingly, the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting.

Unless defined otherwise, all technical and scientific terms used herein have the
25 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 &
30 Marham, The Harper Collins Dictionary of Biology (1991). As used herein, the following terms have the meanings ascribed to them below, unless specified otherwise.

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. 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. For example, “about” can mean within 1 or more than 1 standard deviation, per the practice in the art. Alternatively, “about” can mean a range of up to 20%, up to 10%, up to 5%, or up to 1% of a given value. Alternatively, particularly with respect to biological systems or processes, the term can mean within an order of magnitude, *e.g.*, within 5-fold, within 2-fold of a value. Where particular values are described in the application and claims, unless otherwise stated, the term “about” means within an acceptable error range for the particular value should be assumed.

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.

By “adenosine deaminase” is meant a polypeptide or fragment thereof capable of catalyzing the hydrolytic deamination of adenine or adenosine. 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, such as a bacterium.

In some embodiments, the deaminase or deaminase domain is a variant of a naturally-occurring deaminase from an organism, such as a human, chimpanzee, gorilla, monkey, cow, dog, rat, or mouse. In some embodiments, the deaminase or deaminase domain does not occur in nature. For example, in some embodiments, the deaminase or deaminase domain is 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 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a naturally-occurring deaminase. In some embodiments, the adenosine deaminase is from a bacterium, such as, *E. coli*, *S. aureus*, *S. typhi*, *S. putrefaciens*, *H. influenzae*, or *C. crescentus*. In some embodiments, the adenosine deaminase is a TadA deaminase. In some embodiments, the TadA deaminase is an *E. coli* TadA (ecTadA) deaminase or a fragment thereof.

For example, deaminase domains 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); 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:eaa04774 (2017)), and Rees, H.A., *et al.*, “Base editing: precision chemistry on the genome and transcriptome of living cells.” *Nat Rev Genet.* 2018 Dec;19(12):770-788. doi: 10.1038/s41576-018-0059-1, the entire contents of which are hereby incorporated by reference.

A wild type TadA(wt) adenosine deaminase has the following sequence (also termed TadA reference sequence):

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHP
GMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSD.

In some embodiments, the adenosine deaminase comprises an alteration in the following sequence:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
 GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

(also termed TadA*7.10).

5 The present invention features novel nucleobase editors, where the alterations are made relative to a TadA*7.10 reference sequence.

 In some embodiments, TadA*7.10 comprises at least one alteration. In some embodiments, TadA*7.10 comprises an alteration at amino acid 82 and/or 166. In particular
 10 embodiments, a variant of the above-referenced sequence comprises one or more of the following alterations: Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R. The alteration Y123H refers to the alteration H123Y in TadA*7.10 reverted back to Y123H TadA(wt). In other embodiments, a variant of the TadA*7.10 sequence comprises one or more of the following alterations R21N, R23H, E25F, N38G, L51W, P54C, M70V, Q71M, N72K, Y73S, M94V, P124W, T133K, D139L, D139M, C146R, and A158K of SEQ ID NO:

15 1. In some embodiments, a variant of the TadA*7.10 sequence comprises a combination of alterations selected from the group consisting 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 +
 20 Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R.

 In other embodiments, the invention provides adenosine deaminase variants that include deletions, *e.g.*, TadA*8, comprising a deletion of the C-terminus beginning at residue
 25 149, 150, 151, 152, 153, 154, 155, 156, or 157, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

 In still other embodiments, the adenosine deaminase variant is a homodimer comprising two adenosine deaminase domains each having one or more of the following alterations Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R relative to
 30 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, the adenosine deaminase variant is a heterodimer comprising a wild-type TadA 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 comprising a wild-type TadA 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, the adenosine deaminase variant is a heterodimer comprising 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 comprising a TadA*7.10 domain and an adenosine deaminase variant domain (e.g. TadA*8) comprising a combination of the following alterations: 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; or I76Y + V82S + Y123H + Y147R + Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In one embodiment, the adenosine deaminase is a TadA*8 that comprises or consists essentially of the following sequence or a fragment thereof having adenosine deaminase activity:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD.

In some embodiments, the TadA*8 is truncated. In some embodiments, the truncated
5 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.

10 In particular embodiments, an adenosine deaminase heterodimer comprises an
TadA*8 domain and an adenosine deaminase domain selected from one of the following:

Staphylococcus aureus (*S. aureus*) TadA:

MGSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAH
AEHIAIERAAKVLGSRLEGCTLYVTLEPCVMCAGTIVMSRIPRVYGADDPKGGCSGS
15 LMNLLQQSNFNHRAIVDKGVLKEACSTLLTTFKNLRANKKSTN

Bacillus subtilis (*B. subtilis*) TadA:

MTQDELYMKEAIKEAKKAEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEML
VIDEACKALGTWRLEGATLYVTLEPCPMCAGAVVLSRVEKVVFGAFDPKGGCSGTLMN
LLQEERFNHQAEEVSVGVLEEECGGMLSAFFRELKRRKKKAARKNLSE

20 *Salmonella typhimurium* (*S. typhimurium*) TadA:

MPPAFITGVTSLSDVELDHEYWMRHALTLAKRAWDEREVPVGAVLVHNHRVIGEG
WNRPIGRHDPTAHAEIMALRQGGLVLQNYRLDITLYVTLEPCVMCAGAMVHSRIG
RVVFGARDAKTGAAGSLIDVLHHPGMNHRVEIIEGVLRDECATLLSDFFRMRRQEIK
ALKKADRAEGAGPAV

25 *Shewanella putrefaciens* (*S. putrefaciens*) TadA:

MDEYWMQVAMQMAEKAEAAAGEVPVGAVLVKDGQQIATGYNLSISQHDPTAHAEI
LCLRSAGKKLENYRLLDATLYITILEPCAMCAGAMVHSRIARVVYGARDEKTGAAGT
VVNLLQHAPAFNHQVEVTSGVLAEACSAQLSRFFKRRRDEKKALKLAQRAQQGIE

Haemophilus influenzae F3031 (*H. influenzae*) TadA:

30 MDAAKVRSEFDEKMMRYALELADKAELGEIPVGAVLVDDARNIIGEGWNLSIVQSDPTAH
AEIIALRNGAKNIQNYRLNSTLYVTLEPCTMCAGAILHSRIKRLVFGASDYK
TGAIGSRFHFFDDYKMNHTLEITSGVLAEECQKLSTFFQKRREEKKIEKALLKSLSDK

Caulobacter crescentus (*C. crescentus*) TadA:

MRTDESEDQDHRMMRLALDAARAAAEAGETPVGAVILDPSTGEVIATAGNGPIAAH
 DPTAHAEIAAMRAAAAKLGNRYRLTDLTLVVLTLEPCAMCAGAI SHARIGRVVFGADD
 PKGGAVVHGPKEFAQPTCHWRPEVTGGVLADESADLLRGFFRARRKAKI

Geobacter sulfurreducens (*G. sulfurreducens*) TadA:

5 MSSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRHNLRGSSN
 DPSAHAEMIAIRQAARRSANWRLTGATLYVTLEPCLMCMGAIILARLERVVFGCYDP
 KGGAGSLYDLSADPRLNHQVRLSPGVCQEECGTMLSDFFRDLRRRKKAKATPALF
 IDERKVPPEP

TadA*7.10

10 MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
 GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD.

By “Adenosine Deaminase Base Editor 8 (ABE8) polynucleotide” is meant a polynucleotide encoding an ABE8.

15 By “Adenosine Deaminase Base Editor 9 (ABE9) polypeptide” or “ABE9” is meant a base editor as defined herein comprising an adenosine deaminase variant (TadA*9) comprising one or more alterations at positions ssssss of the sequence shown below. In an embodiment, the adenosine deaminase variant (TadA*9) comprises following alterations: R21N, R23H, E25F, N38G, L51W, P54C, M70V, Q71M, N72K, Y73S, V82T, M94V, P124W, T133K,
 20 D139L, D139M, C146R, and A158K, in the following reference sequence:

	10	20	30	40	50
	MSEVEFSHEY	WMRHALTLAK	<u>RARDE</u> REVPV	GAVLVLN <u>NRV</u>	IGEGWNRAIG
	60	70	80	90	100
	<u>LHDPT</u> AHAEI	MALRQGGLV <u>M</u>	<u>QNY</u> RLIDATL	<u>YV</u> TFEPCVMC	AGAM <u>I</u> HSRIG
25	110	120	130	140	150
	RVVFGVRNAK	TGAAGSLMDV	LHY <u>P</u> GMNHRV	EIT <u>E</u> EGILADE	CAALL <u>C</u> YFFR
	160				
	MPRQVFNA <u>A</u> QK	KAQSSTD.			

30 The relevant bases altered in the reference sequence are shown by underlining and bold font. In some embodiments, ABE9 comprises further alterations, as described herein, relative to the reference sequence.

By “Adenosine Deaminase Base Editor 9 (ABE9) polynucleotide” is meant a polynucleotide encoding an ABE9.

35 By “alpha-1 antitrypsin (A1AT) protein” is meant a polypeptide or fragment thereof having at least about 95% amino acid sequence identity to UniProt Accession No. P01009. In particular embodiments, an A1AT protein comprises one or more alterations relative to the

following reference sequence. In one particular embodiment, an A1AT protein associated with A1AD comprises an E342K mutation. An exemplary A1AT amino acid sequence is >sp|P01009|A1AT_HUMAN Alpha-1-antitrypsin OS=Homo sapiens OX=9606 GN=SERPINA1 PE=1 SV=3, having the following amino acid sequence:

5 MPSSVSWGILLLAGLCCLVPVSLAEDPQGDAAQKTDTSHHQDHPTFNKITPNLAEFASFSLY
 RQLAHQSNSTNIFFS PVS IATAFAMLSLGTKADTHDEILEGLNFNLTETPEAQIHEGFQELL
 RTLNQPDSQLQLTTGNGLFLSEGLKLVDFLEDVKKLYHSEAFTVNFGDTEEAKQINDYVE
 KGTQGKIVDLVKELDRD TVFALVNYIFFK GKWERPF EVKDTEEDDFHVDQVTTVKVPMMKRL
 GMFNIQHCKKLSSWVLLMKYLG NATAIFFLPDEGKLQHLNENLTHDIITKFLENEDRRSASL
 10 HLPKLSITGT YDLKSVLGQLGITKVF SNGADLSGVTEEAPLKLSKAVHKAVLTIDEKGTEAA
 GAMFLEAIPMSIPPEVKFNKPFVFLMIEQNTKSP LFMGKVVNPTQK. In this A1AT protein
 sequence, the first 24 amino acids constitute the signal peptide (underlined). Position 342 of
 the sequence, which is mutated in A1AD (i.e., E342K), is determined based on setting amino
 acid residue “E” following the signal sequence as amino acid “1”.

15 “Administering” is referred to herein as providing one or more compositions
 described herein to a patient or a subject. By way of example and without limitation,
 composition administration, *e.g.*, injection, can be performed by intravenous (i.v.) injection,
 sub-cutaneous (s.c.) injection, intradermal (i.d.) injection, intraperitoneal (i.p.) injection, or
 intramuscular (i.m.) injection. One or more such routes can be employed. Parenteral
 20 administration can be, for example, by bolus injection or by gradual perfusion over time.
 Alternatively, or concurrently, administration can be by an oral route.

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

25 By “alteration” is meant a change (increase or decrease) in the sequence, expression
 levels, or activity of a 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.

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

30 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 (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 in conjunction with a guide polynucleotide (*e.g.*, guide RNA). In various embodiments, the agent is a biomolecular complex comprising a protein domain having base editing activity, *i.e.*, a domain capable of modifying a base (*e.g.*, A, T, C, G, or U) within a nucleic acid molecule (*e.g.*, DNA). In some embodiments, the polynucleotide programmable DNA binding domain is fused or linked to a deaminase domain. In one embodiment, the agent is a fusion protein comprising one or more domains having base editing activity. In another embodiment, the protein domains having base editing activity are linked to the guide RNA (*e.g.*, via an RNA binding motif on the guide RNA and an RNA binding domain fused to the deaminase). In some embodiments, the domains having base editing activity are capable of deaminating a base within a nucleic acid molecule. In some embodiments, the base editor is capable of deaminating one or more bases within a DNA molecule. In some embodiments, the base editor is capable of deaminating a cytosine (C) or an adenosine (A) within DNA. In some embodiments, the base editor is capable of deaminating a cytosine (C) and an adenosine (A) within DNA. In some embodiments, the base editor is a cytidine base editor (CBE). In some embodiments, the base editor is an adenosine base editor (ABE). In some embodiments, the base editor is an adenosine base editor (ABE) and a cytidine base editor (CBE). In some embodiments, the base editor is a nuclease-inactive Cas9 (dCas9) fused to an adenosine deaminase. In some embodiments, the Cas9 is a circular permutant Cas9 (*e.g.*, spCas9 or saCas9). Circular permutant Cas9s are known in the art and described, for example, in Oakes et al., Cell 176, 254–267, 2019. In some embodiments, the base editor is fused to an inhibitor of base excision repair, for example, a UGI domain, or a dISN domain. In some embodiments, the fusion protein comprises a Cas9 nickase fused to a deaminase and an inhibitor of base excision repair, such as a UGI or dISN domain. In other embodiments the base editor is an abasic base editor.

In some embodiments, an adenosine deaminase is evolved from TadA. In some embodiments, the polynucleotide programmable DNA binding domain is a CRISPR

associated (*e.g.*, Cas or Cpf1) enzyme. In some embodiments, the base editor is a catalytically dead Cas9 (dCas9) fused to a deaminase domain. In some embodiments, the base editor is a Cas9 nickase (nCas9) fused to a deaminase domain. In some embodiments, the base editor is fused to an inhibitor of base excision repair (BER). In some embodiments, the inhibitor of base excision repair is a uracil DNA glycosylase inhibitor (UGI). In some
 5 embodiments, the inhibitor of base excision repair is an inosine base excision repair inhibitor. Details of base editors 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.*,

10 “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); 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*
 15 3:eaao4774 (2017), and Rees, H.A., *et al.*, “Base editing: precision chemistry on the genome and transcriptome of living cells.” *Nat Rev Genet.* 2018 Dec;19(12):770-788. doi: 10.1038/s41576-018-0059-1, the entire contents of which are hereby incorporated by reference.

In some embodiments, base editors are generated (*e.g.*, ABE8 or ABE9) by cloning an
 20 adenosine deaminase variant (*e.g.*, TadA*8) into a scaffold that includes a circular permutant Cas9 (*e.g.*, spCAS9) and a bipartite nuclear localization sequence. Circular permutant Cas9s are known in the art and described, for example, in Oakes *et al.*, *Cell* 176, 254–267, 2019. Exemplary circular permutant sequences are set forth below, in which the bold sequence indicates sequence derived from Cas9, the italics sequence denotes a linker sequence, and the
 25 underlined sequence denotes a bipartite nuclear localization sequence.

CP5 (with MSP “NGC=Pam Variant with mutations Regular Cas9 likes NGG” PID=Protein Interacting Domain and “D10A” nickase):

EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
 SM
 PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVAKVEK
 30 GKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKR
 M
 LASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISE
 FSKRVILADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYR
 STKEVL DATLIHQSI TGLYETRIDLSQLGGDGGSGGSGGSGGSGGSGGSGGMDKKYSIGLAI
 GTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT

RRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
 YHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLF E
 ENPINASGVDAKAILSARLSKSRRL ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLA
 EDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASM
 5 IKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKM
 DGTEELLV KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPFLKDNREKIEKILT
 FRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDKNLPNEKV
 LPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
 KKIECFDSVEISGVEDRFNASLGTYHDLKI IKDKDFLDNEENEDI EDIVLTTLTFEDREM
 10 IEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
 MQLIHDDSLTFKEDIQKAQVSGQGD SLHEHIANLAGSPA I KKGILQTVKVVDELVKVMGRHK
 PENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
 NGRDMYVDQELDINRLSDYDV DHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNPSEEVVKM
 KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
 15 KYDENDKLIREVKVITLKS KLVSDFRKDFQFYK VREINNYHHAHDAYLNAVVG TALIKKYPK
 LESEFVYGDYKVYDVRKMI AKSEQEGADKRTADGSEFESPKKKRKV*

In some embodiments, the ABE8 is selected from a base editor from Table 10, 11 or
 13 *infra*. In some embodiments, ABE8 contains an adenosine deaminase variant evolved
 from TadA. In some embodiments, the adenosine deaminase variant of ABE8 is a TadA*8
 20 variant as described in Table 8, 10, 11, or 13 *infra*. In some embodiments, the adenosine
 deaminase variant is the TadA*7.10 variant (*e.g.*, TadA*8) comprising one or more of an
 alteration selected from the group consisting of Y147T, Y147R, Q154S, Y123H, V82S,
 T166R, and/or Q154R. In various embodiments, ABE8 comprises TadA*7.10 variant (*e.g.*
 TadA*8) with a combination of alterations selected from the group of Y147T + Q154R;
 25 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.

30 In some embodiments, the ABE8 is a monomeric construct containing one copy of a
 TadA deaminase, *e.g.*, a TadA*8 variant. In some embodiments, the ABE8 is a dimeric or
 heterodimeric construct containing more than one, *e.g.*, two, copies of the same or different
 TadA deaminase, *e.g.*, a wild-type TadA and a TadA*8 variant.

In some embodiments, the ABE9 is selected from a base editor from Table 14 *infra*. In some embodiments, ABE9 contains an adenosine deaminase variant evolved from TadA. In some embodiments, the adenosine deaminase variant of ABE9 is a TadA*7.10 variant as described in Table 14. In some embodiments, the adenosine deaminase variant is TadA*7.10 comprising one or more alterations selected from the group consisting of Y147T, Y147R, Q154S, Y123H, V82S, T166R, Q154R. In various embodiments, ABE9 comprises TadA*7.10 with alterations selected from the following: Y147R + Q154R + Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y147T + Q154R; Y147T + Q154S; V82S + Q154S; V82T + Q154S and Y123H + Y147R + Q154R + I76Y, in addition to those listed in Table 14. In some embodiments, the ABE9 is a monomeric construct containing one copy of a TadA deaminase, e.g., a TadA*9 variant. In some embodiments, the ABE9 is a dimeric or heterodimeric construct containing more than one, e.g., two, copies of the same or different TadA deaminase, e.g., a wild-type TadA and a TadA*9 variant.

In some embodiments the ABE9 base editor comprises the sequence:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAFLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD.

By way of example, the adenine base editor ABE to be used in the base editing compositions, systems and methods described herein has the nucleic acid sequence (8877 base pairs), (Addgene, Watertown, MA.; Gaudelli NM, *et al.*, Nature. 2017 Nov 23;551(7681):464-471. doi: 10.1038/nature24644; Koblan LW, *et al.*, Nat Biotechnol. 2018 Oct;36(9):843-846. doi: 10.1038/nbt.4172.) as provided below. Polynucleotide sequences having at least 95% or greater identity to the ABE nucleic acid sequence are also encompassed.

ATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACAT
GACCTTATGGGACTTTCCTACTTGCCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGG
TTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTG
ACGTCAATGGGAGTTTGTGGTGGCACC AAAATCAACGGGACTTTCCAAAATGTCGTAACAACCTCCGCCCC
ATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGT
CAGATCCGCTAGAGATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACCATGAAACGGACA
GCCGACGGAAGCGAGTTCGAGTCACCAAAGAAGAAGCGGAAAGTCTCTGAAGTCGAGTTTAGCCACGAGT
ATTGGATGAGGCACGCACTGACCCTGGCAAAGCGAGCATGGGATGAAAGAGAAGTCCCCGTGGGCGCCGT
GCTGGTGCACAACAATAGAGTGATCGGAGAGGGATGGAACAGGCCAATCGGCCGCCACGACCCTACCGCA
CACGCAGAGATCATGGCACTGAGGCAGGGAGGCCTGGTCATGCAGAATTACCGCCTGATCGATGCCACCC
TGTATGTGACACTGGAGCCATGCGTGATGTGCGCAGGAGCAATGATCCACAGCAGGATCGGAAGAGTGGT
GTTCCGAGCACGGGACGCCAAGACCGGCGCAGCAGGCTCCCTGATGGATGTGCTGCACCACCCGGCATG

AACCACCGGTGGAGATCACAGAGGGAATCCTGGCAGACGAGTGCGCCGCCCTGCTGAGCGATTTCTTTA
 GAATGCGGAGACAGGAGATCAAGGCCGAGAAGAAGGCACAGAGCTCCACCGACTCTGGAGGATCTAGCGG
 AGGATCCTCTGGAAGCGAGACACCAGGCACAAGCGAGTCCGCCACACCAGAGAGCTCCGGCGGCTCCTCC
 GGAGGATCCTCTGAGGTGGAGTTTTCCACGAGTACTGGATGAGACATGCCCTGACCCTGGCCAAGAGGG
 5 CACGCGATGAGAGGGAGGTGCCTGTGGGAGCCGTGCTGGTGTGAACAATAGAGTGATCGGCGAGGGCTG
 GAACAGAGCCATCGGCCTGCACGACCCAACAGCCCATGCCGAAATTATGGCCCTGAGACAGGGCGGCCTG
 GTCATGCAGAACTACAGACTGATTGACGCCACCCTGTACGTGACATTGAGCCTTGCGTGATGTGCGCCG
 GCGCCATGATCCACTCTAGGATCGGCCGCGTGGTGTGTTGGCGTGAGGAACGAAAAACGGCGCCGAGG
 CTCCCTGATGGACGTGCTGCACTACCCCGGCATGAATCACCGCGTCGAAATTACCGAGGGAATCCTGGCA
 10 GATGAATGTGCCGCCCTGCTGTGCTATTTCTTTTCGGATGCCTAGACAGGTGTTCAATGCTCAGAAGAAGG
 CCCAGAGCTCCACCGACTCCGGAGGATCTAGCGGAGGCTCCTCTGGCTCTGAGACACCTGGCACAAGCGA
 GAGCGCAACACCTGAAAGCAGCGGGGGCAGCAGCGGGGGTCTCAGACAAGAAGTACAGCATCGGCCTGGCC
 ATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATTCAGG
 TGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTGACAGCGGCGA
 15 AACAGCCGAGGCCACCCGGCTGAAGAGAACC GCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGC
 TATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGT
 CCTTCTGTGTTGAAGAGGATAAGAAGCAGAGCGGCACCCCATCTTCGGCAACATCGTGACGAGGTGGC
 CTACCACGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAAGGCCGAC
 CTGCGGCTGATCTATCTGGCCCTGGCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACC
 20 TGAACCCCGACAACAGCGACGTGGACAAGCTGTTTATCCAGCTGGTGCAGACCTACAACCAGCTGTTTCA
 GGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGA
 CGGCTGGAAAATCTGATCGCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTTCGGAAACCTGATTGCC
 TGAGCCTGGGCCTGACCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAACTGCAGCTGAG
 CAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCAGATCGGCGACCACTACGCCGACCTGTTT
 25 CTGGCCGCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCA
 AGGCCCCCTGAGCGCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGC
 TCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCC
 GGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGG
 ACGGCACCGAGGAACCTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAA
 30 CGGCAGCATCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTAC
 CCATTCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCGCATCCCTACTACGTGGGCC
 CTCTGGCCAGGGGAAACAGCAGATTCTGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCTGGAA
 CTTTCGAGGAAGTGGTGGACAAGGGCGCTTCGCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAG
 AACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGC
 35 TGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGC
 CATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAG
 AAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACAT
 ACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGA
 AGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCC
 40 CACCTGTTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCC
 GGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAAGTCCGACGG

CTTTCGCCAACAGAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAA
 GCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTA
 AGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGA
 GAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGA
 5 ATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGAAAACA
 CCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGA
 ACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGAC
 TCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCAGGGGCAAGAGCGACAACGTGCCCTCCGAAG
 AGGTCTGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTT
 10 CGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAG
 CTGGTGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACG
 ACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCCTGAAGTCCAAGCTGGTGTCCGATTTCCG
 GAAGGATTTCCAGTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAAC
 GCCGTCTGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACA
 15 AGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTT
 CTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGG
 CCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGC
 GGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAA
 AGAGTCTATCCTGCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAG
 20 TACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGAAAAGGGCAAGT
 CCAAGAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAA
 TCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAG
 TACTCCCTGTTTCGAGCTGGAACCGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAA
 ACGAACTGGCCCTGCCCTCCAAATATGTGAACCTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGG
 25 CTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAAGCACTACCTGGACGAGATCATC
 GAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCT
 ACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAA
 TCTGGGAGCCCCCTGCCGCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAA
 GAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTC
 30 AGCTGGGAGGTGACTCTGGCGGCTCAAAAAGAACCGCCGACGGCAGCGAATTCGAGCCCAAGAAGAAGAG
 GAAAGTCTAACCGGTCTATCATCACCATCACCATTGAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTT
 CTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCGTGCCTTCTTACCCCTGGAAGGTGCCACTCCCAC
 TGTCTTTTCTAATAAAATGAGGAAATTGCATCGCATTTGTCTGAGTAGGTGTCATTCTATTCTGGGGGT
 GGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCT
 35 CTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCGATAACCGTCGACCTCTAGCTAGAGCTTGGCGTA
 ATCATGGTCATAGCTGTTTCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGA
 AGCATAAAGTGTAAGCCTAGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGC
 CCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGGGGAGAGGCGG
 TTTGCGTATTGGGCGCTCTTCCGCTTCTCGCTCACTGACTCGCTGCGCTCGGTCTTCCGCTGCGGCGA
 40 GCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACA
 TGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCT

CCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAA
 AGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCCTCTCCTGTTCCGACCCCTGCCGCTTACCGGAT
 ACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTC
 GGTGTAGGTGCTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTACAGCCGACCGCTGCGCCTTA
 5 TCCGGTAAGTATCGTCTTGAGTCCAACCCGGTAAGACAGACTTATCGCCACTGGCAGCAGCCACTGGTA
 ACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTA
 CACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGC
 TCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGAAGCAGCAGATTACGCGCA
 GAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACACTCAGTGGAAACGAAAACTC
 10 ACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGA
 AGTTTTAAATCAATCTAAAGTATATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGG
 CACCTATCTCAGCGATCTGTCTATTTCTGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTAC
 GATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCA
 GATTTATCAGCAATAAACAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCTGCAACTTTATCCGCCT
 15 CCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTGCCAGTTAATAGTTTGCACAACGT
 TGTGTCATGCTACAGGCATCGTGGTGTACGCTCGTCTTGGTATGGCTTCATTCAGCTCCGGTTCC
 CAACGATCAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAAGCGGTTAGCTCCTTCGGTCTCCGA
 TCGTTGTCAGAAGTAAGTTGGCCGCAAGTGTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTAC
 TGTGATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGT
 20 ATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAA
 AAGTGCTCATCATTGAAAAACGTTCTTCGGGGCGAAAACCTCTCAAGGATCTTACCGCTGTTGAGATCCAG
 TTCGATGTAACCCACTCGTGACCCCACTGATCTTCAGCATCTTTTACTTTTACCAGCGTTTCTGGGTGA
 GCAAAAACAGGAAGGCAAAATGCCGCAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATAC
 TCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATG
 25 TATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTGACGGA
 TCGGGAGATCGATCTCCCGATCCCCTAGGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAA
 GCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAAC
 AAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGAT
 GTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCAT
 30 TAGTTCATAGCCCATATATGGAGTTCGCGGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCC
 CAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCAT
 TGACGTCAATGGGTGGAGTATTTACGGTAACTGCCCACTTGGCAGTACATCAAGTGTATC

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
 35 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. In another embodiment, the base editing
 activity is cytidine deaminase activity, *e.g.*, converting target C•G to T•A and adenosine or
 adenine deaminase activity, *e.g.*, converting A•T to G•C.

The term “base editor system” refers to a system 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., a 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 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 casn1 nuclease or a CRISPR (clustered regularly interspaced short palindromic repeat) associated nuclease. An exemplary Cas9, is

Streptococcus pyogenes Cas9 (spCas9), the amino acid sequence of which is provided below:

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETA~~EAT~~
 RLKRTARRRYTRRKNRICYLQEIFS~~NEMAKVDD~~SEFFHRLEESFLVEEDKKHERHPIFGNIVD
 EVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFI
 QLVQIYNQLFEENPINASRVDAKAILSARLSKSRLENLIAQLPGEKRNGLFGNLIALSLGL
 TPNFKNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNS
 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
 YKFIKPILEKMDGTEELLVKLNRDLLRKQRTFDNGSIHQIHLGELHAILRRQEDFYFPLK
 DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMT
 NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIV

LTTLTLFEDRGMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
 LKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIV
DELVKVMGHKPENIVEMARENQTTQKGQNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
 NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF⁵IKDDSIDNKVLTRSDKNRGKSDN
 VPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV
AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHADAYLNAV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANG
EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSD
 KLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGSKKLKSVKELLGITIMERSSFEKNPI
 10 DFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
 YEKLKGSPEQNEQKQLFVEQHKHYLDEIIIEQISEFSKRVILADANLDKVL SAYNKH¹⁰HRDKPIR
 EQAENIIHLFTLTNLGAPAAF¹⁰KYFDTTIDRKRYTSTKEVL¹⁰DATLIHQSI¹⁰TGLYETRIDLSQL
 GGD (single underline: HNH domain; double underline: RuvC domain)

The term “Cas12b” or “Cas12b domain” refers to an RNA-guided nuclease
 15 comprising a Cas12b/C2c1 protein, or a fragment thereof (*e.g.*, a protein comprising an
 active, inactive, or partially active DNA cleavage domain of Cas12b, and/or the gRNA
 binding domain of Cas12b). contents of each of which are incorporated herein by reference).
 Cas12b orthologs have been described in various species, including, but not limited to,
Alicyclobacillus acidoterrestris, *Alicyclobacillus acidophilus* (Teng *et al.*, Cell Discov. 2018
 20 Nov 27;4:63), *Bacillus hisashi*, and *Bacillus sp. V3-13*. Additional suitable Cas12b nucleases
 and sequences will be apparent to those of skill in the art based on this disclosure.

In some embodiments, proteins comprising Cas12b or fragments thereof are referred
 to as “Cas12b variants.” A Cas12b variant shares homology to Cas12b, or a fragment thereof.
 For example, a Cas12b variant is at least about 70% identical, at least about 80% identical, at
 25 least about 90% identical, at least about 95% identical, at least about 96% identical, at least
 about 97% identical, at least about 98% identical, at least about 99% identical, at least about
 99.5% identical, or at least about 99.9% identical to wild type Cas12b. In some
 embodiments, the Cas12b variant may have 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,
 30 42, 43, 44, 45, 46, 47, 48, 49, 50 or more amino acid changes compared to wild type Cas12b.
 In some embodiments, the Cas12b variant comprises a fragment of Cas12b (*e.g.*, a gRNA
 binding domain or a DNA-cleavage domain), such that the fragment is at least about 70%
 identical, at least about 80% identical, at least about 90% identical, at least about 95%
 identical, at least about 96% identical, at least about 97% identical, at least about 98%

identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to the corresponding fragment of wild type Cas12b. In some embodiments, the fragment is 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%,
 5 at least 95% identical, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% of the amino acid length of a corresponding wild type Cas12b. Exemplary Cas12b polypeptides are listed below.

Cas12b/C2c1 (uniprot.org/uniprot/T0D7A2#2) sp|T0D7A2|C2C1_ALIAG CRISPR-
 10 associated endonuclease C2c1 OS = *Alicyclobacillus acido- terrestris* (strain ATCC 49025 / DSM 3922/ CIP 106132 / NCIMB 13137/GD3B) GN=c2c1 PE=1 SV=1
 MAVKSIKVKLRLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQEC
 KTAECKAEELLERLRARQVENGHGPGAGSDDELLQLARQLYELLVPQAIGAKGDAQQIARKE
 LSPLADKDAVGGGLGIAKAGNKPRWVRMREAGEPGWEEEEKKAETRSADRTADVLRLADFG
 15 LKPLMRVYTDSEMSVVEWKPLRKGQAVRTWDRDMFQQAIERMMSWESWNQRVGQEYAKLVEQ
 KNRFEQKNFVGQEHVLVNLQQLQDMKEASPGLESKEQTAHYVTGRALRGSDKVFKEWGLA
 PDAPFDLYDAEIKNVQRRNTRRFGSHDLFAKLAEPEYQALWREDASFLTRYAVYNSILRKLN
 HAKMFATFTLPDATAHPWIWTRFDKLGGLHGYTFLENEFGERRHAIRFHKLLKVENGVAEVE
 DDVTVPISMSEQLDNLPRDPNEPIALYFRDYGAEQHFTGEFGGAKIQCRDQLAHMHRRRG
 20 ARDVYLNVSVRVQSQSEARGERRPPYAAVFRVLDGNHRAVHFDKLSDYLAEHPDDGKLGSE
 GLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNKGRVPFFFPKGNLNLVAVHERSOLL
 KLPGETESKDLRAIREERQRTLRLQLRTQLAYLRLLVRCGSEDVGRRRERSWAKLIEQPVDAA
 HMTPDWREAFENELQKLKSLHGICSDKEWMDAVYESVRRVVRHMGKQVRDWRKDVRSGERPK
 IRGYAKDVVGNSIEQIEYLERQYKFLKSWSFFGKVSQVIRAEKGSRFAITLREHIDHAKE
 25 DRLKKLADRIIMEALGYVYALDERGKGKWKVAKYPPCQLILLEELSEYQFNDRPPSENNQLM
 QWSHRGVFQELINQAQVHDLVGTMYAAFFSRFDARTGAPGIRCRVPPARCTQEHNPFPFPW
 WLNKFVVEHTLDACPLRADDLIPTGEGEIFVSPFSAEEGDFHQIHADLNAAQNLQQLWSDF
 DISQIRLRCDWGEVDGELVLIPLRTGKRTADSYSNKVFTNTGVTTYERERGGKRRKVFAQE
 KLSEEEAELLVEADEAREKSVVLMRDPGSIINRGNWTRQKEFWSMVNQRIEGLVKQIRSRV
 30 PLQDSACENTGDI

AacCas12b (*Alicyclobacillus acidiphilus*) - WP_067623834

MAVKSMKVKLRLDNMPEIRAGLWKLHTEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQECY
 KTAECKAEELLERLRARQVENGHGCGPAGSDDELLQLARQLYELLVPQAIGAKGDAQQIARKE

LSPLADKDAVGGLGIAKAGNKPRWVRMREAGEPGWEEEEKAKAEARKSTDRTADVLRALADFG
 LKPLMRVYTDSMSSVQWKPLRKGQAVRTWDRDMFQQAIERMMSWESWNQRVGEAYAKLVEQ
 KSRFEQKNFVQGHEHLVQLVNQLQQDMKEASHGLESKEQTAHYLTGRALRGSDKVFKEKWEKLD
 PDAPFDLYDTEIKNVQRRNTRRFGSHDLFAKLAEPKYQALWREDASFLTRYAVYNSIVRKLN
 5 HAKMFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGEGRHAI RFQKLLTVEDGVAKEV
 DDVTVPISMSAQLDDLLPRDPHELVALYFQDYGAEQHLAGEFGGAKIQYRRDQLNHLHARRG
 ARDVYLNLSVRVQSQSEARGERRPPYAAVFRLVGDNHRA FVHFDKLSDYLA EHPDDGKLGSE
 GLLSGLRVMSVDLGLRTSASISVFRVAR KDELKPNSEGRVPFCFPIEGNENLVAVHERS QLL
 KLPGETESKDLRAIREERQRTLRLQLRTQLAYLRLLVRCGSEDVGRRERSWAKLIEQPM DANQ
 10 MTPDWREAFED ELQKLKSLYGICGDREWTEAVYESVRRVVRHMGKQVRDWRKDVRSGERP KI
 RGYQKDVVGGNSIEQIEYLERQYKFLKSWSFFGKVSGQVIRAEKGS RFAITLREHIDHAKED
 RLKKLADRIIMEALGYVYALDDER GKGKWKVAKYPPCQLILLEELSEYQFNND RPPSENNQLM
 QWSHRGVFQELLNQAQVHDL LVGTMYAAFSSRFDARTGAPGIRCR RVPARCAREQNPEFPW
 WLNKFVAEHKLDGCPLRADDLIPTGEGEFFVSPFSAEEGDFHQIHADLNAAQNLQRRLWSD F
 15 DISQIRLRCDWGEVDGEPVLI PRTTGKRTADSYGNKV FYTKTGVTYYERER GKKRRKVFAQE
 ELSEEEAELLVEADEAREKSVVLMRDP SGIINRGDWTRQKEFWSMVNQRIEGYLVKQIRSRV
 RLQESACENTGDI

BhCas12b (*Bacillus hisashii*) NCBI Reference Sequence: WP_095142515

20 MAPKKKRKVG IHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAIYYMNI LKLI RQEAI
 YEHHEQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVE
 KKGEANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEE EKKKWEEDKKKDP
 LAKILGKLAEYGLIPLFI PYTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWES
 WNLKVKEEYKVEKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLR
 25 GWREIIQKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEF LSKKENHFIWRNHPEYPY
 LYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRI LTEQLHTEKLKKKL
 TVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDES IKFPLKGT
 LGGARVQFDRDHLRRYPHKVESGNVGRIYFNMVTNIEPTESPVSKSLKIHRDDFPKV VNFKP
 KELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGRQAAAASIFEVVDQKPDIEGKLFFPIK
 30 GTELYAVHRASFNIKLPGETLVKSREVLRKAREDNLKL MNQKLNFLRNVLHFQQFEDITERE
 KRVTKWISRQENS DVPLVYQDELIQIRELMYKPYKDWVAF LKQLHKRLEVEIGKEVKHWRKS
 LSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKED
 RLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYPY EERSRFENSKLM KWSR
 REIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCS **S**VVTKEKLQDNRRFFKNLQREGR

LTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTTHADINAAQNLQKRFWTRTHGIFYKVYCK
 AYQVDGQTVYIPESKDQKQKIIIEFEGGYFILKDGVEWVNAGKLKIKKGSSKQSSSELVDS
 DILKDSFDLASELKGEKLMLYRDPGSGNVFSPDKWMAAGVFFGKLERILISKLTNQYSISTIE
 DDSSKQSMKRPAATKKAGQAKKKK

- 5 The variant termed BvCas12b V4 includes the changes S893R, K846R, and E837G relative to the wild-type sequence above.

BvCas12b (*Bacillus* sp. V3-13) NCBI Reference Sequence: WP_101661451.1

MAIRSIKLMKMTNSGTDSIYLRKALWRTHQLINEGIAYYMNLLTLRQEAIGDKTKE
 10 AYQAELINIIRNQQRNNGSSEEHGSDQEILALLRQLYELIIPSSIGESGDANQLGNKFLYPL
 VDPNSQSGKGTSNAGRKPRWKRLKEEGNPDWELEKKKDEERKAKDPTVKIFDNLNKYGLLPL
 FPLFTNIQKDIEWLPLGKRQSVRKWDKDMFIQAIERLLSWESWNRVADYEYKQLEKTESYY
 KEHLTGGEEWIEKIRKFEKERNMELEKNAFAPNDGYFITSRQIRGWDRVYEKWSKLPEASAP
 EELWKVVAEQQNMSEGFGDPKVFSFLANRENDRDIWRGHSEIRYHIAAYNGLQKKLSRTKEQ
 15 ATFTLPDAIEHPLWIRYESPGGTNLNLFKLEEKQKKNYVTLISKIIPSEEKWIENIEIP
 LAPSIQFNRIKQHVKGKQEISFSDYSSRISLDGVLGGSRIQFNRYIKNHKELLGEGDI
 GPVFFNLVVDVAPLQETRNGRLQSPIGKALKVISSDFSVIDYKPKELMDWMNTGSASNSFG
 VASLLEGMRVMSIDMGQRTSASVSIFEVVKELPKDQEQKLFYSINDTELFAIHKRSFLLNLP
 GEVVTKNKQQRQERRKKRQFVRSQIRMLANVLRLETCKTPDERKKAIHKLMEIVQSYDSWT
 20 ASQKEVWEKELNLLTNMAAFNDEIWKESLVELHHRIEYPYVGQIVSKWRKGLSEGRKNLAGIS
 MWNIDELEDTRRLNISWSKRSRTPGEANRIETDEPFGSSLLQHIQNVKDDRLKQMANLIIMT
 ALGFKYDKEEKDRYKRWKETYPACQIILFENLNRYLFNLDNRSRRENSRLMKWAHRSIPRTVS
 MQGEMFGLQVGDVRSSEYSSRFHAKTGAPGIRCHALTEEDLKAGSNTLKRLIEDGFINESELA
 YLKKGDIIIPSQGGELFVTLISKRYKKDSNNELTVIHADINAAQNLQKRFWQQNSEVYRVPCQ
 25 LARMGEDKLYIPKSQTETIKKYFGKGSFVKNNTEQEVYKWEKSEMKIKTDTTDFLDQDLGDF
 EDISKTIELAQQKKYLTMFRDPSGYFFNNETWRPQKEYWSIVNNIIKSLCKKKILSNKVE
 L.

- 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
 30 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 –NH₂ 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 CAG → TAG Stop codon

CAA → TAA

Arginine CGA → TGA

Tryptophan TGG → TGA

TGG → TAG

TGG → TAA

By “cytidine deaminase” is meant a polypeptide or fragment thereof capable of catalyzing a deamination reaction that converts an amino group to a carbonyl group. In one embodiment, the cytidine deaminase converts cytosine to uracil or 5-methylcytosine to thymine. PmCDA1, which is derived from *Petromyzon marinus* (*Petromyzon marinus* cytosine deaminase 1, “PmCDA1”), AID (Activation-induced cytidine deaminase; AICDA), which is derived from a mammal (*e.g.*, human, swine, bovine, horse, monkey etc.), and APOBEC are exemplary cytidine deaminases.

The term “deaminase” or “deaminase domain,” as used herein, refers to a protein or enzyme that catalyzes a deamination reaction. In some embodiments, the deaminase or deaminase domain is a cytidine deaminase, catalyzing the hydrolytic deamination of cytidine or deoxycytidine to uridine or deoxyuridine, respectively. In some embodiments, the deaminase or deaminase domain is a cytosine deaminase, catalyzing the hydrolytic deamination of cytosine to uracil. In some embodiments, the deaminase is an adenosine deaminase, which catalyzes the hydrolytic deamination of adenosine to inosine. In some embodiments, the deaminase is an adenosine deaminase, which catalyzes the hydrolytic deamination of adenosine or adenine (A) to inosine (I). In some embodiments, the deaminase

or deaminase domain is an adenosine deaminase, catalyzing the hydrolytic deamination of adenosine or deoxyadenosine to inosine or deoxyinosine, respectively. In some embodiments, the adenosine deaminase catalyzes the hydrolytic deamination of adenosine in deoxyribonucleic acid (DNA). The adenosine deaminase (*e.g.*, engineered adenosine deaminase, evolved adenosine deaminase) provided herein can be from any organism, such as a bacterium. In some embodiments, the adenosine deaminase is from a bacterium, such as *E. coli*, *S. aureus*, *S. typhi*, *S. putrefaciens*, *H. influenzae*, or *C. crescentus*. In some embodiments, the adenosine deaminase is a TadA deaminase. In some embodiments, the deaminase or deaminase domain is a variant of a naturally occurring deaminase from an organism, such as a human, chimpanzee, gorilla, monkey, cow, dog, rat, or mouse. In some embodiments, the deaminase or deaminase domain does not occur in nature. For example, in some embodiments, the deaminase or deaminase domain is 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 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%, at least 99.1%, at least 99.2%, at least 99.3%, at least 99.4%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, or at least 99.9% identical to a naturally occurring deaminase.

“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.

In some embodiments, an effective amount of a fusion protein provided herein, *e.g.*, of a nucleobase editor comprising a nCas9 domain and a deaminase domain (*e.g.*, adenosine deaminase, cytidine deaminase) refers to the amount that is sufficient to induce editing of a target site specifically bound and edited by the nucleobase editors described herein. As will be appreciated by the skilled artisan, the effective amount of an agent, *e.g.*, a fusion protein, may vary depending on various factors as, for example, on the desired biological response, *e.g.*, on the specific allele, genome, or target site to be edited, on the cell or tissue being targeted, and/or on the agent being used.

In some embodiments, an effective amount of a fusion protein provided herein, *e.g.*, of a fusion protein comprising a nCas9 domain and a deaminase domain may refer to the amount of the fusion protein that is sufficient to induce editing of a target site specifically bound and edited by the fusion protein. As will be appreciated by the skilled artisan, the effective amount of an agent, *e.g.*, a fusion protein, a nuclease, a hybrid protein, a protein dimer, a complex of a protein (or protein dimer) and a polynucleotide, or a polynucleotide, may vary depending on various factors as, for example, on the desired biological response, *e.g.*, on the specific allele, genome, or target site to be edited, on the cell or tissue being targeted, and/or on the agent being used.

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 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 RNA” or “gRNA” is meant a polynucleotide that 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. gRNAs that exist as a single RNA molecule may be referred to as single-guide RNAs (sgRNAs), although “gRNA” is used interchangeably to refer to guide RNAs that exist as either single molecules or as a complex of two or more molecules. Typically, gRNAs that exist as single RNA species comprise two domains: (1) a domain that shares homology to a target nucleic acid (e.g., and directs binding of a Cas9 complex to the target); and (2) a domain that binds a Cas9 protein. In some embodiments, domain (2) corresponds to a sequence known as a tracrRNA, and comprises a stem-loop structure. For example, in some embodiments, domain (2) is identical or homologous to a tracrRNA as provided in Jinek et al., Science 337:816-821(2012), the entire contents of which is incorporated herein by reference. Other examples of gRNAs (e.g., those including domain 2) can be found in US20160208288, entitled "Switchable Cas9 Nucleases and Uses Thereof," and US 9,737,604, entitled "Delivery System For Functional Nucleases," the entire contents of each are hereby incorporated by reference in their entirety. In some embodiments, a gRNA comprises two or more of domains (1) and (2), and may be referred to as an “extended gRNA.” An extended gRNA will bind two or more Cas9 proteins and bind a target nucleic acid at two or more distinct regions, as described herein. The gRNA comprises a nucleotide sequence that complements a target site, which mediates binding of the nuclease/RNA complex to the target site, providing the sequence specificity of the nuclease:RNA complex.

By “heterodimer” is meant a fusion protein comprising two domains, such as a wild type TadA domain and a variant of TadA domain (e.g., TadA*8 or TadA*9) or two variant TadA domains (e.g., TadA*7.10 and TadA*8 or two TadA*8 domains; or TadA*7.10 and TadA*9 or two TadA*9 domains).

"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. In some embodiments, the IBR is an inhibitor of inosine base excision repair. Exemplary inhibitors of base repair

5 include inhibitors of APE1, Endo III, Endo IV, Endo V, Endo VIII, Fpg, hOGG1, hNEIL1, T7 EndoI, T4PDG, UDG, hSMUG1, and hAAG. In some embodiments, the base repair inhibitor is an inhibitor of Endo V or hAAG. In some embodiments, the IBR is an inhibitor of Endo V or hAAG. In some embodiments, the IBR is a catalytically inactive EndoV or a catalytically inactive hAAG. In some embodiments, the base repair inhibitor is a catalytically inactive

10 EndoV or a catalytically inactive hAAG. In some embodiments, the base repair inhibitor is uracil glycosylase inhibitor (UGI). UGI refers to a protein that is capable of inhibiting a uracil-DNA glycosylase base-excision repair enzyme. In some embodiments, a UGI domain comprises a wild-type UGI or a fragment of a wild-type UGI. In some embodiments, the UGI proteins provided herein include fragments of UGI and proteins homologous to a UGI or

15 a UGI fragment. In some embodiments, the base repair inhibitor is an inhibitor of inosine base excision repair. In some embodiments, the base repair inhibitor is a “catalytically inactive inosine specific nuclease” or “dead inosine specific nuclease.” Without wishing to be bound by any particular theory, catalytically inactive inosine glycosylases (*e.g.*, alkyl adenine glycosylase (AAG)) can bind inosine, but cannot create an abasic site or remove the

20 inosine, thereby sterically blocking the newly formed inosine moiety from DNA damage/repair mechanisms. In some embodiments, the catalytically inactive inosine specific nuclease can be capable of binding an inosine in a nucleic acid but does not cleave the nucleic acid. Non-limiting exemplary catalytically inactive inosine specific nucleases include catalytically inactive alkyl adenosine glycosylase (AAG nuclease), for example, from a

25 human, and catalytically inactive endonuclease V (EndoV nuclease), for example, from *E. coli*. In some embodiments, the catalytically inactive AAG nuclease comprises an E125Q mutation or a corresponding mutation in another AAG nuclease.

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.

30 Inteins are also referred to as “protein introns.” The process of an intein excising itself and joining the remaining portions of the protein is herein termed “protein splicing” or “intein-mediated protein splicing.” In some embodiments, an intein of a precursor protein (an intein containing protein prior to intein-mediated protein splicing) comes from two genes. Such

intein is referred to herein as a split intein (e.g., split intein-N and split intein-C). For example, in cyanobacteria, DnaE, the catalytic subunit a of DNA polymerase III, is encoded by two separate genes, dnaE-n and dnaE-c. The intein encoded by the dnaE-n gene may be herein referred as "intein-N." The intein encoded by the dnaE-c gene may be herein referred as "intein-C."

Other intein systems may also be used. For example, 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). Non-limiting examples of intein pairs 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.

DnaE Intein-N DNA:

TGCCTGTCATACGAAACCGAGATACTGACAGTAGAATATGGCCTTCTGCCAATCGGGAAGAT
TGTGGAGAAACGGATAGAATGCACAGTTTACTCTGTGCGATAACAATGGTAACATTTATACTC
AGCCAGTTGCCAGTGGCACGACCGGGGAGAGCAGGAAGTATTCGAATACTGTCTGGAGGAT
GGAAGTCTCATTAGGGCCACTAAGGACCACAAATTTATGACAGTCGATGGCCAGATGCTGCC
TATAGACGAAATCTTTGAGCGAGAGTTGGACCTCATGCGAGTTGACAACCTTCCTAAT

DnaE Intein-N Protein:

CLSYETEILTVEYGLLPKIVEKRIECTVYSVDNNGNIYTQPVAQWHDR
GEQEVFEYCLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNL PN

DnaE Intein-C DNA:

ATGATCAAGATAGCTACAAGGAAGTATCTTGGCAAACAAAACGTTTATGA
TATTGGAGTCGAAAGAGATCACAACCTTTGCTCTGAAGAACGGATTCATAG CTTCTAAT

Intein-C: MIKIATRKYLKGQNVYDIGVERDHNFKNGFIASN

Cfa-N DNA:

TGCCTGTCTTATGATACCGAGATACTTACCGTTGAATATGGCTTCTTGCCTATTGGAAAGAT
TGTCGAAGAGAGAATTGAATGCACAGTATATACTGTAGACAAGAATGGTTTCGTTTACACAC
AGCCCATTGCTCAATGGCACAATCGCGGCGAACAAGAAGTATTTGAGTACTGTCTCGAGGAT
GGAAGCATCATACGAGCAACTAAAGATCATAAATTCATGACCACTGACGGGCAGATGTTGCC
AATAGATGAGATATTCGAGCGGGGCTTGGATCTCAAACAAGTGGATGGATTGCCA

Cfa-N Protein:

CLSYDTEILTVEYGFLLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLED
GSIIRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP

Cfa-C DNA:

ATGAAGAGGACTGCCGATGGATCAGAGTTTGAATCTCCCAAGAAGAAGAGGAAAGTAAAGAT
5 AATATCTCGAAAAAGTCTTGGTACCCAAAATGTCTATGATATTGGAGTGGAGAAAGATCACA
ACTTCCTTCTCAAGAACGGTCTCGTAGCCAGCAAC

Cfa-C Protein:

MKRTADGSEFESPKKKRKVKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN

Intein-N and intein-C may be fused to the N-terminal portion of the split Cas9 and the
10 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
15 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,
20 US20150344549, and US20180127780, each of which is incorporated herein by reference in
their entirety.

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"
25 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,
30 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 (e.g., a DNA) 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. 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, can refer to a covalent linker (e.g., covalent bond), a non-covalent linker, a chemical group, or a molecule linking two molecules or moieties, e.g., two components of a protein complex or a ribonucleocomplex, or two domains of a fusion protein, such as, for example, a polynucleotide programmable DNA binding domain (e.g., dCas9) and a deaminase domain (e.g., an adenosine deaminase, a cytidine deaminase, or an adenosine deaminase and a cytidine deaminase). A linker can join different components of, or different portions of components of, a base editor system. For example, in some embodiments, a linker can join a guide polynucleotide binding domain of a polynucleotide programmable nucleotide binding domain and a catalytic domain of a deaminase. In some embodiments, a linker can join a CRISPR polypeptide and a deaminase. In some embodiments, a linker can join a Cas9 and a deaminase. In some embodiments, a

linker can join a dCas9 and a deaminase. In some embodiments, a linker can join a nCas9 and a deaminase. In some embodiments, a linker can join a guide polynucleotide and a deaminase. In some embodiments, a linker can join a deaminating component and a polynucleotide programmable nucleotide binding component of a base editor system. In some embodiments, a linker can join a RNA-binding portion of a deaminating component and a polynucleotide programmable nucleotide binding component of a base editor system. In some embodiments, a linker can join a RNA-binding portion of a deaminating component and a RNA-binding portion of a polynucleotide programmable nucleotide binding component of a base editor system. A linker can be positioned between, or flanked by, two groups, molecules, or other moieties and connected to each one via a covalent bond or non-covalent interaction, thus connecting the two. In some embodiments, the linker can be an organic molecule, group, polymer, or chemical moiety. In some embodiments, the linker can be a polynucleotide. In some embodiments, the linker can be a DNA linker. In some embodiments, the linker can be a RNA linker. In some embodiments, a linker can comprise an aptamer capable of binding to a ligand. In some embodiments, the ligand may be carbohydrate, a peptide, a protein, or a nucleic acid. In some embodiments, the linker may comprise an aptamer may be derived from a riboswitch. The riboswitch from which the aptamer is derived may be selected from a theophylline riboswitch, a thiamine pyrophosphate (TPP) riboswitch, an adenosine cobalamin (AdoCbl) riboswitch, an S-adenosyl methionine (SAM) riboswitch, an SAH riboswitch, a flavin mononucleotide (FMN) riboswitch, a tetrahydrofolate riboswitch, a lysine riboswitch, a glycine riboswitch, a purine riboswitch, a GlmS riboswitch, or a pre-queosine1 (PreQ1) riboswitch. In some embodiments, a linker may comprise an aptamer bound to a polypeptide or a protein domain, such as a polypeptide ligand. In some embodiments, the polypeptide ligand may be a K Homology (KH) domain, a MS2 coat protein domain, a PP7 coat protein domain, a SfMu Com coat protein domain, a sterile alpha motif, a telomerase Ku binding motif and Ku protein, a telomerase Sm7 binding motif and Sm7 protein, or a RNA recognition motif. In some embodiments, the polypeptide ligand may be a portion of a base editor system component. For example, a nucleobase editing component may comprise a deaminase domain and a RNA recognition motif.

In some embodiments, the linker can be an amino acid or a plurality of amino acids (*e.g.*, a peptide or protein). In some embodiments, the linker can be about 5-100 amino acids in length, for example, about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, or 90-100 amino acids in length. In some

embodiments, the linker can be about 100-150, 150-200, 200-250, 250-300, 300-350, 350-400, 400-450, or 450-500 amino acids in length. Longer or shorter linkers can also be used. Longer or shorter linkers are also contemplated. In some embodiments, a linker comprises the amino acid sequence SGSETPGTSESATPES, which may also be referred to as the XTEN linker. In some embodiments, a linker comprises the amino acid sequence SGGS. In some embodiments, a linker comprises (SGGS)_n, (GGGS)_n, (GGGGS)_n, (G)_n, (EAAAK)_n, (GGS)_n, SGSETPGTSESATPES, or (XP)_n motif, or a combination of any of these, where n is independently an integer between 1 and 30, and where X is any amino acid. In some embodiments, n is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15. 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, PAPAPA, PAPAPAP, PAPAPAPA, P(AP)₄, P(AP)₇, P(AP)₁₀. Such proline-rich linkers are also termed “rigid” linkers.

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 (*e.g.*, cytidine or adenosine deaminase). In some embodiments, a linker joins a dCas9 and a nucleic-acid editing protein. For example, the linker is positioned between, or flanked by, two groups, molecules, or other moieties and connected to each one via a covalent bond, thus connecting the two. In some embodiments, the linker is an amino acid or a plurality of amino acids (*e.g.*, a peptide or protein). In some embodiments, the linker is an organic molecule, group, polymer, or chemical moiety. In some embodiments, the linker is 5-200 amino acids in length, for example, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 35, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 101, 102, 103, 104, 105, 110, 120, 130, 140, 150, 160, 175, 180, 190, or 200 amino acids in length.

In some embodiments, the domains of a base editor are fused via a linker that comprises the amino acid sequence of SGGSSGGSSGSETPGTSESATPESGGSS, SGGSSGGSSGSETPGTSESATPESGGSSGGSS, or GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS. In some embodiments, domains of the base editor are fused via a linker comprising the amino acid sequence SGSETPGTSESATPES, which may also be referred to as the XTEN linker. In some embodiments, the linker is 24 amino acids in length. In some embodiments, the linker comprises the amino acid sequence SGGSSGGSSGSETPGTSESATPES. In some embodiments, the linker is 40 amino acids in length. In some embodiments, the linker

comprises the amino acid sequence

SGGSSGGSSGSETPGTSESATPESGGSSGGSSGGSSGGS. In some embodiments, the linker is 64 amino acids in length. In some embodiments, the linker comprises the amino acid sequence

5 SGGSSGGSSGSETPGTSESATPESGGSSGGSSGGSSGSETPGTSESATPESGGSSGGS. In some embodiments, the linker is 92 amino acids in length. In some embodiments, the linker comprises the amino acid sequence
PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAP
GTSTEPSEGSAPGTSESATPESGPGSEPATS.

10 By “marker” is meant any protein or polynucleotide having an alteration in expression level or activity that is associated with a disease or disorder.

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
15 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)). In some
20 embodiments, the presently disclosed base editors can efficiently generate 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. In some embodiments, an intended mutation is a mutation that is generated by a specific base editor (*e.g.*, cytidine base editor or adenosine base editor) bound
25 to a guide polynucleotide (*e.g.*, gRNA), specifically designed to generate the intended mutation.

In general, mutations made or identified in a sequence (*e.g.*, an amino acid sequence as described herein) are numbered in relation to a reference (or wild type) sequence, *i.e.*, a sequence that does not contain the mutations. The skilled practitioner in the art would readily
30 understand how to determine the position of mutations in amino acid and nucleic acid sequences relative to a reference sequence.

The term “non-conservative mutations” involve amino acid substitutions between different groups, for example, lysine for tryptophan, or phenylalanine for serine, etc. In this

case, it is preferable for the non-conservative amino acid substitution to not interfere with, or inhibit the biological activity of, the functional variant. The non-conservative amino acid substitution can enhance the biological activity of the functional variant, such that the biological activity of the functional variant is increased as compared to the wild-type protein.

5 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
10 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, KRPAATKKAGQAKKKK, KKTELQTTNAENKTKKL, KRGINDRNFWRGENGRKTR, RKSGKIAAIVVKRPRK,
15 PKKKRKV, or MDSLLMNRRKFLYQFKNVRWAKGRRETYLC.

 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)
20 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-
25 dihydrouracil, 5-methylcytosine (m5C), and 5-hydrmethylcytosine. 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
30 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.

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 "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
 10 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
 15 domain, for example a nuclease active Cas9, a Cas9 nickase (nC9), 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/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, and Cas12j/CasΦ. Non-limiting examples of Cas enzymes include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5d, Cas5t,
 20 Cas5h, Cas5a, Cas6, Cas7, Cas8, Cas8a, Cas8b, Cas8c, Cas9 (also known as Csn1 or Csx12), Cas10, Cas10d, Cas12a/Cpf1, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, Cas12j/CasΦ, 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,
 25 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.,
 30 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.

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). In some embodiments, the nucleobase editing domain is more than one deaminase domain (*e.g.*, an adenine deaminase or an adenosine deaminase and a cytidine or a cytosine deaminase). In some embodiments, the nucleobase editing domain can be a naturally occurring nucleobase editing domain. In some embodiments, the nucleobase editing domain can be an engineered or evolved nucleobase editing domain from the naturally occurring nucleobase editing domain. The nucleobase editing domain can be from any organism, such as a bacterium, human, chimpanzee, gorilla, monkey, cow, dog, rat, or mouse.

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 mammals 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 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.

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. The terms refer to a protein, peptide, or polypeptide of any size, structure, or function. Typically, a protein, peptide, or polypeptide will be at least three amino acids long. A protein, peptide, or polypeptide can refer to an individual protein or a collection of proteins. One or more of the amino acids in a protein, peptide, or polypeptide can be modified, for example, by the addition of a chemical entity such as a carbohydrate group, a hydroxyl group, a phosphate group, a farnesyl group, an isofarnesyl group, a fatty acid group, a linker for conjugation, functionalization, or other modifications, etc. A protein, peptide, or polypeptide can also be a single molecule or can be a multi-molecular complex. A protein, peptide, or polypeptide can be just a fragment of a naturally occurring protein or peptide. 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. One protein can be located at the amino-terminal (N-terminal) portion of the fusion protein or at the carboxy-terminal (C-terminal) protein thus forming an amino-terminal fusion protein or a carboxy-terminal fusion protein, respectively. A protein can comprise different domains, for example, a nucleic acid binding domain (*e.g.*, the gRNA binding domain of Cas9 that directs the binding of the protein to a target site) and a nucleic acid cleavage domain, or a catalytic domain of a nucleic acid editing protein. In some embodiments, a protein comprises a proteinaceous part, *e.g.*, an amino acid sequence constituting a nucleic acid binding domain, and an organic compound, *e.g.*, a compound that can act as a nucleic acid cleavage agent. In some embodiments, a protein is in a complex with, or is in association with, a nucleic acid, *e.g.*, RNA or DNA. Any of the proteins provided herein can be produced by any method known in the art. For example, the proteins provided herein can be produced via recombinant protein expression and purification, which is especially suited for fusion proteins comprising a peptide linker. Methods for recombinant protein expression and purification are well known, and include those described by Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)), the entire contents of which are incorporated herein by reference.

Polypeptides and proteins disclosed herein (including functional portions and functional variants thereof) can comprise synthetic amino acids in place of one or more naturally-occurring amino acids. Such synthetic amino acids are known in the art, and include, for example, aminocyclohexane carboxylic acid, norleucine, α -amino n-decanoic acid, homoserine, S-acetylaminoethyl-cysteine, trans-3- and trans-4-hydroxyproline, 4-aminophenylalanine, 4-nitrophenylalanine, 4-chlorophenylalanine, 4-carboxyphenylalanine, β -phenylserine β -hydroxyphenylalanine, phenylglycine, α -naphthylalanine, cyclohexylalanine, cyclohexylglycine, indoline-2-carboxylic acid, 1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid, aminomalonic acid, aminomalonic acid monoamide, N'-benzyl-N'-methyl-lysine, N',N'-dibenzyl-lysine, 6-hydroxylysine, ornithine, α -aminocyclopentane carboxylic acid, α -aminocyclohexane carboxylic acid, α -aminocycloheptane carboxylic acid, α -(2-amino-2-norbornane)-carboxylic acid, α,γ -diaminobutyric acid, α,β -diaminopropionic acid, homophenylalanine, and α -tert-butylglycine. The polypeptides and proteins can be associated with post-translational modifications of one or more amino acids of the polypeptide constructs. Non-limiting examples of post-translational modifications include phosphorylation, acylation including acetylation and formylation, glycosylation (including N-linked and O-linked), amidation, hydroxylation, alkylation including methylation and ethylation, ubiquitylation, addition of pyrrolidone carboxylic acid, formation of disulfide bridges, sulfation, myristoylation, palmitoylation, isoprenylation, farnesylation, geranylation, glypiation, lipoylation and iodination.

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 a wild-type or healthy cell. In other embodiments and without limitation, a reference is 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
 5 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
 10 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 term "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
 15 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 (CsnI) from *Streptococcus pyogenes* (See, e.g., "Complete genome sequence of an M1 strain of
 20 *Streptococcus pyogenes*." Ferretti J.J. 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).

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

systems. *Nature biotechnology* 31, 233-239 (2013); the entire contents of each of which are incorporated herein by reference).

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%). For example, at a specific base position in the human genome, the C nucleotide can appear in most individuals, but in a minority of individuals, the position is occupied by an A. This means that there is a SNP at this specific position, and the two possible nucleotide variations, C or A, are said to be alleles for this position. SNPs underlie differences in susceptibility to disease. The severity of illness and the way our body responds to treatments are also manifestations of genetic variations. SNPs can fall within coding regions of genes, non-coding regions of genes, or in the intergenic regions (regions between genes). In some embodiments, SNPs within a coding sequence do not necessarily change the amino acid sequence of the protein that is produced, due to degeneracy of the genetic code. SNPs in the coding region are of two types: synonymous and nonsynonymous SNPs. Synonymous SNPs do not affect the protein sequence, while nonsynonymous SNPs change the amino acid sequence of protein. The nonsynonymous SNPs are of two types: missense and nonsense. SNPs that are not in protein-coding regions can still affect gene splicing, transcription factor binding, messenger RNA degradation, or the sequence of noncoding RNA. Gene expression affected by this type of SNP is referred to as an eSNP (expression SNP) and can be upstream or downstream from the gene. A single nucleotide variant (SNV) is a variation in a single nucleotide without any limitations of frequency and can arise in somatic cells. A somatic single nucleotide variation can also be called a single-nucleotide alteration.

By “specifically binds” is meant a nucleic acid molecule, polypeptide, or complex thereof (*e.g.*, a nucleic acid programmable DNA binding domain and guide nucleic acid), 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.

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.

5 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)

10 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.,

15 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

20 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

30 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. In particular embodiments, the Cas9 protein is divided into two fragments within a disordered region of the protein, e.g., as described in Nishimasu et al., Cell, Volume 156, Issue 5, pp. 935-949, 2014, or as described in Jiang et al. (2016) Science 351: 867-871. PDB file: 5F9R, each of which is incorporated herein by reference. In some embodiments, the protein is divided into two fragments at any C, T, A, or S within a region of SpCas9 between about amino acids A292-G364, F445-K483, or E565-T637, or at corresponding positions in any other Cas9, Cas9 variant (e.g., nCas9, dCas9), or other napDNAbp. In some embodiments, protein is divided into two fragments at SpCas9 T310, T313, A456, S469, or C574. In some embodiments, the process of dividing the protein into two fragments is referred to as “splitting” the protein.

In other embodiments, the N-terminal portion of the Cas9 protein comprises amino acids 1-573 or 1-637 S. pyogenes Cas9 wild-type (SpCas9) (NCBI Reference Sequence: NC_002737.2, Uniprot Reference Sequence: Q99ZW2) and the C-terminal portion of the Cas9 protein comprises a portion of amino acids 574-1368 or 638-1368 of SpCas9 wild-type.

The C-terminal portion of the split Cas9 can be joined with the N-terminal portion of the split Cas9 to form a complete Cas9 protein. In some embodiments, the C-terminal portion of the Cas9 protein starts from where the N-terminal portion of the Cas9 protein ends. As such, in some embodiments, the C-terminal portion of the split Cas9 comprises a portion of amino acids (551-651)-1368 of spCas9. "(551-651)-1368" means starting at an amino acid between amino acids 551-651 (inclusive) and ending at amino acid 1368. For example, the C-terminal portion of the split Cas9 may comprise a portion of any one of amino acid 551-1368, 552-1368, 553-1368, 554-1368, 555-1368, 556-1368, 557-1368, 558-1368, 559-1368, 560-1368, 561-1368, 562-1368, 563-1368, 564-1368, 565-1368, 566-1368, 567-1368, 568-1368, 569-1368, 570-1368, 571-1368, 572-1368, 573-1368, 574-1368, 575-1368, 576-1368, 577-1368, 578-1368, 579-1368, 580-1368, 581-1368, 582-1368, 583-1368, 584-1368, 585-1368, 586-1368, 587-1368, 588-1368, 589-1368, 590-1368, 591-1368, 592-1368, 593-1368, 594-1368, 595-1368, 596-1368, 597-1368, 598-1368, 599-1368, 600-1368, 601-1368, 602-1368, 603-1368, 604-1368, 605-1368, 606-1368, 607-1368, 608-1368, 609-1368, 610-1368, 611-1368, 612-1368, 613-1368, 614-1368, 615-1368, 616-1368, 617-1368, 618-1368, 619-1368, 620-1368, 621-1368, 622-1368, 623-1368, 624-1368, 625-1368, 626-1368, 627-1368, 628-1368, 629-1368, 630-1368, 631-1368, 632-1368, 633-1368, 634-1368, 635-1368, 636-1368, 637-1368, 638-1368, 639-1368, 640-1368, 641-1368, 642-1368, 643-1368, 644-1368, 645-1368, 646-1368, 647-1368, 648-1368, 649-1368, 650-1368, or 651-1368 of spCas9. In some embodiments, the C-terminal portion of the split Cas9 protein comprises a portion of amino acids 574-1368 or 638-1368 of SpCas9.

By "subject" is meant a mammal, including, but not limited to, a human or non-human mammal, such as a non-human primate (monkey), bovine, equine, canine, ovine, or feline. In some embodiments, a subject described herein includes a pathogenic mutation in a polynucleotide sequence.

By "substantially identical" is meant a polypeptide or nucleic acid molecule exhibiting at least 50% identity to a reference amino acid sequence (for example, any one of the amino acid sequences described herein) or nucleic acid sequence (for example, any one of the nucleic acid sequences described herein). In one embodiment, such a sequence is at least 60%, 80% or 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, COBALT, EMBOSS Needle, 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
 5 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.

10 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
 15 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;
- 20 d) OUTPUT FORMAT: pair;
- e) END GAP PENALTY: false;
- f) END GAP OPEN: 10; and
- g) END GAP EXTEND: 0.5.

The term "target site" refers to a sequence within a nucleic acid molecule that is
 25 deaminated by a deaminase (e.g., cytidine or adenine deaminase) or a fusion protein comprising a deaminase (e.g., a dCas9-adenosine deaminase fusion protein or a base editor disclosed herein).

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
 30 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. In one embodiment, the agent is a protein or fragment thereof that binds a host uracil-DNA glycosylase and prevents removal of uracil residues from DNA. In an embodiment, a UGI is a protein, a fragment thereof, or a domain that is capable of inhibiting a uracil-DNA glycosylase base-excision repair enzyme. In some embodiments, a UGI domain comprises a wild-type UGI or a modified version thereof. In some embodiments, a UGI domain comprises a fragment of the exemplary amino acid sequence set forth below. In some embodiments, a UGI fragment comprises an amino acid sequence that comprises 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 100% of the exemplary UGI sequence provided below. In some embodiments, a UGI comprises an amino acid sequence that is homologous to the exemplary UGI amino acid sequence or fragment thereof, as set forth below. In some embodiments, the UGI, or a portion thereof, is 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%, at least 99.5%, at least 99.9%, or 100% identical to a wild type UGI or a UGI sequence, or portion thereof, as set forth below. An exemplary UGI comprises an amino acid sequence as follows:

>splP14739IUNGI_BPPB2 Uracil-DNA glycosylase inhibitor

MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSD
APEYKPWALVIQDSNGENKIKML.

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.

Any compositions or methods provided herein can be combined with one or more of any of the other compositions and methods provided herein.

5 The description and examples herein illustrate embodiments of the present disclosure in detail. It is to be understood that this disclosure is not limited to the particular embodiments described herein and as such can vary. Those of skill in the art will recognize that there are numerous variations and modifications of this disclosure, which are encompassed within its scope.

10 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.

15 The practice of some embodiments disclosed herein employ, unless otherwise indicated, conventional techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics and recombinant DNA, which are within the skill of the art. See for example Sambrook and Green, *Molecular Cloning: A Laboratory Manual*, 4th Edition (2012); the series *Current Protocols in Molecular Biology* (F. M. Ausubel, *et al.* eds.); the series *Methods In Enzymology* (Academic Press, Inc.), PCR 2: A
20 Practical Approach (M.J. MacPherson, B.D. Hames and G.R. Taylor eds. (1995)), Harlow and Lane, eds. (1988) *Antibodies, A Laboratory Manual*, and *Culture of Animal Cells: A Manual of Basic Technique and Specialized Applications*, 6th Edition (R.I. Freshney, ed. (2010)).

25 Although various features of the present disclosure can be described in the context of a single embodiment, the features can also be provided separately or in any suitable combination. Conversely, although the present disclosure can be described herein in the context of separate embodiments for clarity, the present disclosure can also be implemented in a single embodiment. The section headings used herein are for organizational purposes only and are not to be construed as limiting the subject matter described.

30 The features of the present disclosure are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in

which the principles of the disclosure are utilized, and in view of the accompanying drawings as described hereinbelow.

BRIEF DESCRIPTION OF THE DRAWINGS

5 **FIG. 1** presents a series of graphs showing percent A>G editing activity for the designated adenosine base editors. Each of the editors is referred to by number where, for example, 433 denotes pNMG-B433, which is ABE8.32. Each of the editors referenced in the graph was tested with each of gRNAs HRB03, HRB04, HRB08, HRB12, and ng-424. The gRNA sequences are provided in Example 3.

10 **FIG. 2** provides a heat map depicting in gray shading percent A>G editing activity for the designated adenosine base editors (ABE8 and ABE9), which are described at Table 14. Each of the editors listed in the figure was tested with a different gRNA, HRB03, HRB04, HRB08, HRB12, and ng-424.

FIGS. 3A-3C provide tables showing TadA deaminase variant (e.g., TadA*9; ABE9) and Cas9 (e.g., SpCas9) variant components of adenosine base editors described herein. These ABE9 base editors have A>G editing activity and are useful for correcting SNP mutations associated with alpha-1 antitrypsin disease (A1AD), such as the PiZ mutation in the *SERPINA1* gene. In some cases, the SpCas9 variants have specificity for 5'-NGC-3' PAMs. **FIG. 3A** refers to the adenosine base editors by their plasmid number. **FIGS. 3B and 3C** present various TadA deaminase variants and amino acid mutations included in the Tad*7.10 amino acid sequence, as well as PAM variants and their included amino acid mutations.

FIGS. 4A-4D present a nucleic acid sequence, a table and graphs related to producing improved rates of nucleobase correction through base editor engineering. **FIGS. 4A and 4B** present a nucleic acid sequence and a table related to producing improved rates of nucleobase correction in primary PiZZ fibroblasts through base editor engineering as described in **FIGS. 4C and 4D** and related to increasing serum alpha-1 antitrypsin (A1AT) produced by lipid nanoparticle (LNP)-mediated delivery and base editing in NSG-PiZ transgenic mice as described in **FIGS. 5A and 5B** *infra*. In particular, **FIG. 4A** shows the target DNA sequence, including the target site (the A at position 7 in the target DNA sequence), encoding the PiZZ mutation associated with A1AD. This sequence includes the 20 nucleotide protospacer and a non-canonical spCas9 NGC PAM. Shown also are beneficial edits at position A7 = wild-type (WT) and edits at positions A5 and A7 = WT + D341G. **FIG. 4B**

presents a table describing the TadA deaminase variant and the Cas9 PAM variant constituents of the various base editors used to correct the PiZ mutation. The table shows the variants (e.g., Variants (Vars) 1-9) as used to obtain the results provided in **FIGS. 4C, 4D, 5A and 5B**. In the table, amino acid mutations in SpCas9 (SpCas9 variants) are depicted in the rightmost column of the table (PAM variant). The “RVRFRAR” SpCas9 variant includes the following mutations: L1111R + D1135V + G1218R + E1219F + A1322R + R1335A + T1337R. **FIGS. 4C and 4D** present bar-graphs depicting the editing rates observed in patient-derived PiZZ fibroblasts (GM1423 Corriel Biorepository) that were transfected with base editing reagents using the Neon electroporation system. Each treatment consisted of 10µl electroporation buffer containing 70,000 fibroblasts, 100ng mRNA encoding the base editor and 50ng Alpha-1 correction gRNA. After 48 hours of recovery, the cells were lysed, and the locus of interest was interrogated by targeted amplicon sequencing. The data were obtained from two independent experiments. These data and results demonstrate the improvements in target base editing efficiency from both optimization of the NGC PAM recognition (variants 1-3, **FIGS. 4B and 4C**) and optimization of the TadA deaminase through incorporation of mutations in the TadA deaminase, e.g., ABE9, (variants 4-9, **FIGS. 4B-4D**).

FIGS. 5A and 5B present graphs related to the increase in serum A1AT produced by lipid nanoparticle (LNP)-mediated delivery and base editing in NSG-PiZ transgenic mice. The target site DNA sequence and the table of the TadA deaminase variant and Cas9 PAM variant constituents of the various editors used to correct the PiZ mutation are as described in **FIGS. 4A and 4B** above. **FIG. 5A** presents a graph depicting the editing rates observed in total liver gDNA from the NSG-PiZ transgenic mouse model 7 days after treatment with 1.5 mg/kg of LNP containing a 1:1 weight ratio of gRNA and mRNA encoding base editor. Commercially available NSG-PiZ mice (The Jackson Laboratory, Mount Desert Island, ME) express mutant human SERPINA1 (Glu342Lys mutation) on the immunodeficient NOD-SCID gamma (NSG) background, which provides a stable background for human hepatocytes after partial hepatectomy. The results demonstrate that the ngcABEvar9 (**FIG. 4B**) yielded higher rates of editing than the earlier version variant 8. **FIG. 5B** presents a graph showing that the editing rates are correlated with an increase in serum Alpha-1 Antitrypsin (A1AT), (post-bleed), relative to pretreatment samples, (pre-bleed), as measured by an MSD Sandwich Immunoassay. Based on these results, base editing with the TadA deaminase variants

described herein is capable of addressing a deficiency of alpha-1 antitrypsin and its potential pulmonary sequelae.

DETAILED DESCRIPTION OF THE INVENTION

The invention features novel adenine base editors (e.g., ABE9) and methods of using these adenosine deaminase variants for editing a target sequence.

NUCLEOBASE EDITOR

Disclosed herein are novel base editors (e.g., ABE8 and ABE9) or nucleobase editors for editing, modifying or altering a target nucleotide sequence of a polynucleotide. In particular, the novel ABE9 base editor and its component adenosine deaminase are described in Tables 14 and 18 *infra*. Described herein is a nucleobase editor or a base editor comprising a polynucleotide programmable nucleotide binding domain and a nucleobase editing domain (e.g., adenosine 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 (*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.

Polynucleotide Programmable Nucleotide Binding Domain

It should be appreciated that polynucleotide programmable nucleotide binding domains can also include nucleic acid programmable proteins that bind RNA. For example, the polynucleotide programmable nucleotide binding domain can be associated with a nucleic acid that guides the polynucleotide programmable nucleotide binding domain to an RNA. Other nucleic acid programmable DNA binding proteins are also within the scope of this disclosure, though they are not specifically listed in this disclosure.

A polynucleotide programmable nucleotide binding domain of a base editor can itself comprise one or more domains. For example, a polynucleotide programmable nucleotide binding domain can comprise 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. Herein the term “exonuclease” refers to a protein or

polypeptide capable of digesting a nucleic acid (*e.g.*, RNA or DNA) from free ends, and 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). In some embodiments, an endonuclease can cleave a single strand of a double-stranded nucleic acid. In some
 5 embodiments, an endonuclease can cleave both strands of a double-stranded nucleic acid molecule. In some embodiments a polynucleotide programmable nucleotide binding domain can be a deoxyribonuclease. In some embodiments a polynucleotide programmable nucleotide binding domain can be a ribonuclease.

In some embodiments, a nuclease domain of a polynucleotide programmable
 10 nucleotide binding domain can cut zero, one, or two strands of a target polynucleotide. In some cases, 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
 15 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
 20 mutation and a histidine at position 840. In such cases, 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
 25 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.

30 The amino acid sequence of an exemplary catalytically active Cas9 is as follows:
 MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
 RLKRTARRRYTRKKNRICYLQEIFSNEMAKVDDSSFFHRLEESFLVEEDKKHERHPPIFGNIVD
 EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFI
 QLVQTYNQLFEEINPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIALLSLGL

TPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
 YKFIKPILEKMDGTEELLVKNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLK
 DNREKIEKILTFRIPIYYVGPLARGNSRFAMTRKSEETITPWNFEVVVDKGASQAQSFIERMT
 5 NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
 LTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
 LKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVV
 DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
 10 QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
 NVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
 VAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
 VGTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
 GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKRNS
 15 DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEFKNP
 IDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
 HYEKLKGSPEQKQLFVEQHKHYLDEIIQI SEFSKRVILADANLDKVL SAYNKH RDKPI
 REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDL SQ
 LGGD.

20 A base editor comprising a polynucleotide programmable nucleotide binding domain
 comprising a nickase domain is thus able to generate a single-strand DNA break (nick) at a
 specific polynucleotide target sequence (*e.g.*, determined by the complementary sequence of
 a bound guide nucleic acid). 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.*,
 25 Cas9-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) can cleave the strand of a DNA molecule which is being targeted for
 editing. In such cases, the non-targeted strand is not cleaved.

30 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.

Also contemplated herein are mutations capable of generating a catalytically dead polynucleotide programmable nucleotide binding domain from a previously functional version of the polynucleotide programmable nucleotide binding domain. For example, in the case of catalytically dead Cas9 (“dCas9”), variants having mutations other than D10A and H840A are provided, which result in nuclease inactivated Cas9. Such mutations, by way of example, include other amino acid substitutions at D10 and H840, or other substitutions within the nuclease domains of Cas9 (*e.g.*, substitutions in the HNH nuclease subdomain and/or the RuvC1 subdomain). Additional suitable nuclease-inactive dCas9 domains can 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).

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 cases, 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.

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., Chylinski K., Fonfara I., Hauer M., Doudna J. A., Charpentier E. *Science* 337:816-821(2012), the entire contents of which is hereby incorporated by reference. Cas9 recognizes a short motif in the CRISPR repeat sequences (the PAM or protospacer adjacent motif) to help distinguish self-versus-non-self.

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 (or polynucleotide, *e.g.*, DNA or RNA) target to be modified. Thus, a skilled artisan can change the genomic or polynucleotide target of the Cas protein by changing the target sequence present in the gRNA. The specificity of the Cas protein is partially determined by how specific the gRNA targeting sequence is for the genomic polynucleotide target sequence compared to the rest of the genome. In an embodiment, the Cas protein is SpCas9.

In some embodiments, the gRNA scaffold sequence is as follows:

GUUUUAGAGCUAGAAAUAGCAAGUUAUUAAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGG
CACCGAGUCGGUGCUUUU.

In some embodiments, the gRNA scaffold sequence is as follows:

5 GUUUUAGAGCUAGAAAUAGCAAGUUAUUAAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGG
CACCGAGUCGGUGCUUUU.

In an embodiment, the terminal uracils (U) of above gRNA scaffolds may optionally comprise "mU*mU*mU*U," which denote 2'OMe and have phosphorothioate linkages.

In an embodiment, the RNA scaffold comprises a stem loop. In an embodiment, the

10 RNA scaffold comprises the nucleic acid sequence:

GUUUUUGUACUCUCAAGAUUUAAGUAAACUGUACAACGAAACUACACAGUUACUAAAUCUU
GCAGAAGCUACAAAGAUAAAGGCUUCAUGCCGAAAUCAACACCCUGUCAUUUUUAUGGCAGGGU
G.

In an embodiment, an *S. pyrogenes* sgRNA scaffold polynucleotide sequence is as

15 follows:

GUUUUAGAGCUAGAAAUAGCAAGUUAUUAAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGG
CACCGAGUCGGUGC.

In an embodiment, an *S. aureus* sgRNA scaffold polynucleotide sequence is as follows:

20 GUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUA
UCUCGUCAACUUGUUGGCGAGA.

In an embodiment, a BhCas12b sgRNA scaffold has the following polynucleotide sequence:

GUUCUGTCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGCAGGGUGUGAGAAACUCCU
25 AUUGCUGGACGAUGUCUCUUACGAGGCAUUAGCAC.

In an embodiment, a BvCas12b sgRNA scaffold has the following polynucleotide sequence:

GACCUAUAGGGUCAAUUGAAUCUGUGCGUGUGCCAUAAAGUAAUUAAAAUUACCCACCACAGG
AGCACCUGAAAACAGGUGCUUGGCAC.

30 In some embodiments, a CRISPR protein-derived domain incorporated into a base editor is an endonuclease (e.g., deoxyribonuclease or ribonuclease) capable of binding a target polynucleotide when in conjunction with a bound guide nucleic acid. In some embodiments, a CRISPR protein-derived domain incorporated into a base editor is a nickase capable of binding a target polynucleotide when in conjunction with a bound guide nucleic

acid. In some embodiments, a CRISPR protein-derived domain incorporated into a base editor is a catalytically dead domain capable of binding a target polynucleotide when in conjunction with a bound guide nucleic acid. In some embodiments, a target polynucleotide bound by a CRISPR protein derived domain of a base editor is DNA. In some embodiments, a target polynucleotide bound by a CRISPR protein-derived domain of a base editor is RNA.

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, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, and Cas12j/CasΦ, CARF, DinG, homologues thereof, or modified versions thereof. An unmodified CRISPR enzyme can have DNA cleavage activity, such as Cas9, which has two functional endonuclease domains: RuvC and HNH. 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. Cas9 can refer to a polypeptide 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 Cas9 polypeptide (*e.g.*, Cas9 from *S. pyogenes*). Cas9 can refer to a polypeptide with at most or at most 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 Cas9 polypeptide (*e.g.*, from *S. pyogenes*). Cas9 can refer to the wild type or a modified form of the Cas9 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 domains of Nucleobase Editors

Cas9 nuclease sequences and structures are well known to those of skill in the art (See, e.g., “Complete genome sequence of an M1 strain of *Streptococcus pyogenes*.” Ferretti et al., *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.

In some aspects, a nucleic acid programmable DNA binding protein (napDNAbp) is a Cas9 domain. Non-limiting, exemplary Cas9 domains are provided herein. The Cas9 domain may be a nuclease active Cas9 domain, a nuclease inactive Cas9 domain, or a Cas9 nickase. In some embodiments, the Cas9 domain is a nuclease active domain. For example, the Cas9 domain may be a Cas9 domain that cuts both strands of a duplexed nucleic acid (e.g., both strands of a duplexed DNA molecule). In some embodiments, the Cas9 domain comprises any one of the amino acid sequences as set forth herein. In some embodiments the

Cas9 domain 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 herein. In some embodiments, the Cas9 domain 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, 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 mutations compared to any one of the amino acid sequences set forth herein. In some embodiments, the Cas9 domain comprises an amino acid sequence that has at least 10, at least 15, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 150, at least 200, at least 250, at least 300, at least 350, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, at least 1000, at least 1100, or at least 1200 identical contiguous amino acid residues as compared to any one of the amino acid sequences set forth herein.

In some embodiments, proteins comprising fragments of Cas9 are provided. For example, in some embodiments, a protein comprises one of two Cas9 domains: (1) the gRNA binding domain of Cas9; or (2) the DNA cleavage domain of Cas9. In some embodiments, proteins comprising Cas9 or fragments thereof are referred to as “Cas9 variants.” A Cas9 variant shares homology to Cas9, or a fragment thereof. For example, a Cas9 variant is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to wild type Cas9. In some embodiments, the Cas9 variant may have 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, 50 or more amino acid changes compared to wild type Cas9. In some embodiments, the Cas9 variant comprises a fragment of Cas9 (*e.g.*, a gRNA binding domain or a DNA-cleavage domain), such that the fragment is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to the corresponding fragment of wild type Cas9. In some embodiments, the fragment is 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% identical, at least 96%, at least 97%, at least 98%, at

least 99%, or at least 99.5% of the amino acid length of a corresponding wild type Cas9. In some embodiments, the fragment is at least 100 amino acids in length. In some embodiments, the fragment is at least 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, or at least 1300 amino acids in length.

In some embodiments, Cas9 fusion proteins as provided herein comprise the full-length amino acid sequence of a Cas9 protein, *e.g.*, one of the Cas9 sequences provided herein. In other embodiments, however, fusion proteins as provided herein do not comprise a full-length Cas9 sequence, but only one or more fragments thereof. Exemplary amino acid sequences of suitable Cas9 domains and Cas9 fragments are provided herein, and additional suitable sequences of Cas9 domains and fragments will be apparent to those of skill in the art.

A Cas9 protein can associate with a guide RNA that guides the Cas9 protein to a specific DNA sequence that has complementary to the guide RNA. In some embodiments, the polynucleotide programmable nucleotide binding domain is a Cas9 domain, for example a nuclease active Cas9, a Cas9 nickase (nCas9), or a nuclease inactive Cas9 (dCas9).

Examples of nucleic acid programmable DNA binding proteins include, without limitation, Cas9 (*e.g.*, dCas9 and nCas9), Cas12a/Cpf1, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, and Cas12j/CasΦ. 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/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, Cas12j/CasΦ, 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.

In some embodiments, wild type Cas9 corresponds to Cas9 from *Streptococcus pyogenes* (NCBI Reference Sequence: NC_017053.1, nucleotide and amino acid sequences as follows).

```
ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGAT
CACTGATGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACA
GTATCAAAAAAATCTTATAGGGGCTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACT
CGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACA
```

GGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGT
 CTTTTTTGGTGGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGAT
 GAAGTTGCTTATCATGAGAAATATCCAACATCTATCATCTGCGAAAAAATGGCAGATTC
 TACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTG
 5 GTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAACTATTTATC
 CAGTTGGTACAAATCTACAATCAATTATTTGAAGAAAACCTATTAACGCAAGTAGAGTAGA
 TGCTAAAGCGATTCTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTC
 AGCTCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGAATCTCATTGCTTTGTCATTGGGATTG
 ACCCCTAATTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGA
 10 TACTTACGATGATGATTTAGATAATTTATTGGCGCAAATTTGGAGATCAATATGCTGATTTGT
 TTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATAGT
 GAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAGCGCTACGATGAACATCATCAAGA
 CTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACCTCCAGAAAAGTATAAAGAAATCTTTT
 TTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTT
 15 TATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACT
 AAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTTCCCATCAAA
 TTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAAAA
 GACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTGAATTCTTATTATGTTGGTCCATT
 GGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCAT
 20 GGAATTTTGAAGAAGTTGTGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACA
 AACTTTGATAAAATCTTCAAATGAAAAAGTACTACCAAACATAGTTTGCTTTATGAGTA
 TTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAGGGAATGCGAAAACCAG
 CATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAACAAATCGAAAA
 GTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGA
 25 AATTTACAGGAGTTGAAGATAGATTTAATGCTTCATTAGGCGCCTACCATGATTTGCTAAAAA
 TTATTAAGATAAAGATTTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTT
 TTAACATTGACCTTATTTGAAGATAGGGGGATGATTGAGGAAAGACTTAAACATATGCTCA
 CCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTT
 TGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTT
 30 TTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGAC
 ATTTAAAGAAGATATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGA
 TTGCTAACTTAGCTGGCAGTCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAATTGTT
 GATGAACCTGGTCAAAGTAATGGGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGA
 AAATCAGACAACCTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAG

GTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAA
 AATGAAAAGCTCTATCTCTATTATCTACAAAATGGAAGAGACATGTATGTGGACCAAGAATT
 AGATATTAATCGTTTAAAGTGATTATGATGTGATCACATTGTTCCACAAAGTTTCATTAAAG
 ACGATTCAATAGACAATAAGGTACTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAAC
 5 GTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAACTATTGGAGACAACCTTCTAAACGCCAA
 GTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAAC
 TTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTG
 GCACAAATTTTGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTTCGAGA
 GGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCT
 10 ATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTT
 GGAAGTGCCTTTGATTAAGAAATATCCAAAACCTTGAATCGGAGTTTGTCTATGGTGATTATAA
 AGTTTATGATGTTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAA
 AATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAATGGA
 GAGATTTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAA
 15 AGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGA
 AAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGAC
 AAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAATATGGTGGTTTTGATAGTCCAAC
 GGTAGCTTATTAGTCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAAT
 CCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAATCCGATT
 20 GACTTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAA
 ATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTAC
 AAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAGTCAT
 TATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAACAATTGTTTGTGGAGCAGCA
 TAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTTAG
 25 CAGATGCCAATTTAGATAAAGTTCTTAGTGATATAACAAACATAGAGACAAACCAATACGT
 GAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTT
 TAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGATG
 CCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTA
 GGAGGTGACTGA

30

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETA
 RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVD
 EVAYHEKYPTIYHLRKKLADSTDKADLRLLIYLAHAMIKFRGHFLIEGDLNPDNSDVKLFI
 QLVQIYNQLFEENPINASRVDAKAILSARLSKSRLENLIAQLPGEKRNGLFGNLIALSLGL

TPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNS
 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
 YKFIKPILEKMDGTEELLVKNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLK
 DNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASQAQSFIERMT
 5 NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIV
 LTLTLFEDRGMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
 LKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIV
DELVKVMGHKPENIVEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
 10 NEKLYLYYLQNGRDMYVDQELDINRLSDYDDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDN
VPSEEVVKMKMNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV
AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHADAYLNAV
GTALIKKYPKLESEFVYGDKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANG
EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSD
 15 KLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELLGITIMERSSFEKNPI
 DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
 YEKLKGSPEQNEQQLFVEQHKHYLDEIIIEQISEFSKRVILADANLDKVLSAYNKHDKPIR
 EQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSIITGLYETRIDLSQL
 GGD

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

In some embodiments, wild type Cas9 corresponds to, or comprises the following nucleotide and/or amino acid sequences:

ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCGTTGGATGGGCTGTCAT
 AACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATT
 25 CGATTAAAAAGAATCTTATCGGTGCCCTCCTATTTCGATAGTGGCGAAACGGCAGAGGCGACT
 CGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACCGAATATGTTACTTACA
 AGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGTTTGAAGAGT
 CCTTCCTTGTCTGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGAT
 GAGGTGGCATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTC
 30 AACTGATAAAGCGGACCTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTG
 GGCACCTTCTCATTGAGGGTGATCTAAATCCGGACAACCTCGGATGTCGACAACTGTTTCATC
 CAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCTATAAATGCAAGTGGCGTGGA
 TGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTGATCGCAC
 AATTACCCGGAGAGAAGAAAAATGGGTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTG

ACACCAAATTTTAAGTCGAACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGA
 CACGTACGATGACGATCTCGACAATCTACTGGCACAAATTGGAGATCAGTATGCGGACTTAT
 TTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCCTATCTGACATACTGAGAGTTAATACT
 GAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTACGATGAACATCACCAAGA
 5 CTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATATTCT
 TTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTC
 TACAAGTTTATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAACAACT
 CAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCGACAACGGTAGCATTCCACATCAAA
 TCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCGTTCTCCTCAA
 10 GACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGGGACCCCT
 GGCCCCGAGGGAACCTCTCGGTTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCAT
 GGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACC
 AACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTA
 TTTCACAGTGTACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAACCCG
 15 CCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAA
 GTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTCGATTCTGTCTGA
 GATCTCCGGGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAAGA
 TAATTAAGATAAGGACTTCCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTG
 TTGACTCTTACCCTCTTTGAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCTCA
 20 CCTGTTTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGAT
 TGTCGCGGAAACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAACTATTCTCGATTTT
 CTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGACTCTTTAAC
 CTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATA
 TTGCGAATCTTGCTGGTTCCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTG
 25 GATGAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAAACATTGTAATCGAGATGGCAGC
 CGAAAATCAAACGACTCAGAAGGGGGCAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAG
 AGGGTATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTG
 CAGAACGAGAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAGGA
 ACTGGACATAAACCGTTTATCTGATTACGACGTCGATCACATTGTACCCCAATCCTTTTTGA
 30 AGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCGAGGGAAAAGTGAC
 AATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAAATGC
 GAAACTGATAACGCAAAGAAAGTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCTG
 AACTTGACAAGGCCGGATTTATTAAACGTGAGCTCGTGGAACCCGCCAAATCACAAAGCAT
 GTTGACACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATTCTG

GGAAGTCAAAGTAATCACTTTAAAGTCAAATTTGGTGTCTGGACTTCAGAAAGGATTTTCAAT
 TCTATAAAGTTAGGGAGATAAATAACTACCACCATGCGCACGACGCTTATCTTAATGCCGTC
 GTAGGGACCGCACTCATTAAGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTA
 CAAAGTTTATGACGTCCGTAAGATGATCGCGAAAAGCGAACAGGAGATAGGCAAGGCTACAG
 5 CCAAATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAATCACTCTGGCAAAC
 GGAGAGATACGCAAACGACCTTTAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGA
 TAAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAA
 AGAAAAGTGAAGTGCAGACCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAGGAATAGT
 GATAAGCTCATCGCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCC
 10 TACAGTTGCCTATTCTGTCCTAGTAGTGCCAAAAGTTGAGAAGGGAAAATCCAAGAACTGA
 AGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAAAGAACCCC
 ATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAGGATCTCATAATTAACTACC
 AAAGTATAGTCTGTTTGAGTTAGAAAATGGCCGAAACGGATGTTGGCTAGCGCCGGAGAGC
 TTCAAAGGGGAACGAACTCGCACTACCGTCTAAATACGTGAATTCCTGTATTTAGCGTCC
 15 CATTACGAGAAGTTGAAAGGTTACCTGAAGATAACGAACAGAAGCAACTTTTTGTTGAGCA
 GCACAAACATTATCTCGACGAAATCATAGAGCAAATTTCCGAATTCAGTAAGAGAGTCATCC
 TAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATAACAAGCACAGGGATAAACCCATA
 CGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGC
 ATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGGAGGTGCTAG
 20 ACGCGACACTGATTCACCAATCCATCACGGGATTATATGAAACTCGGATAGATTTGTCACAG
 CTTGGGGGTGACGGATCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTACAAAGACCATGA
 CGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGGCTGCAGGA

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
 25 RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVD
 EVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHAMIKFRGHFLIEGDLNPDNSDVKLFI
 QLVQTYNQLFEEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALLSLGL
 TPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
 30 YKFIKPILEKMDGTEELLVKNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLLK
 DNREKIEKILTFRIPIYVGPLARGNSRFAMTRKSEETITPWNFEVVDKGASAQSFIERMT
 NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGYHDLLKIIKDKDFLDNEENEDILEDIV
 LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF

LKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVEMARENQTTQKGQNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
 5 VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHADAYLNAV
VGTALIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
 DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELLGITIMERSSEFKNP
 IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
 10 HYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFSKRVILADANLDKVL SAYNKH RDKPI
 REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI TGLYETRIDL SQ
 LGGD

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

In some embodiments, wild type Cas9 corresponds to Cas9 from *Streptococcus*
 15 *pyogenes* (NCBI Reference Sequence: NC_002737.2 (nucleotide sequence as follows); and
 Uniprot Reference Sequence: Q99ZW2 (amino acid sequence as follows):
 ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGAT
 CACTGATGAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACA
 GTATCAAAAAAATCTTATAGGGGCTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCGACT
 20 CGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACA
 GGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGT
 CTTTTTTGGTGGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGAT
 GAAGTTGCTTATCATGAGAAATATCCAATCTATCATCTGCGAAAAAATGGTAGATTC
 TACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTG
 25 GTCATTTTTTGAATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAACTATTTATC
 CAGTTGGTACAAACCTACAATCAATTATTTGAAGAAAACCTATTAACGCAAGTGGAGTAGA
 TGCTAAAGCGATTCTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTC
 AGCTCCCCGGTGAGAAGAAAAATGGCTTATTTGGGAATCTCATTGCTTTGTCATTGGGTTTG
 ACCCCTAATTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAGA
 30 TACTTACGATGATGATTTAGATAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGT
 TTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATACT
 GAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAACGCTACGATGAACATCATCAAGA
 CTTGACTCTTTTAAAGCTTTAGTTCGACAACAACCTCCAGAAAAGTATAAAGAAATCTTTT
 TTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTT

TATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACT
 AAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCAAA
 TTCACCTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTTAAAA
 GACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATT
 5 GGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCAT
 GGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTATTGAACGCATGACA
 AACTTTGATAAAAAATCTTCCAAATGAAAAAGTACTACCAAACATAGTTTGCTTTATGAGTA
 TTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAAGGAATGCGAAAACCAG
 CATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAA
 10 GTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGA
 AATTTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGTACCTACCATGATTGCTAAAAA
 TTATTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTT
 TTAACATTGACCTTATTTGAAGATAGGGAGATGATTGAGGAAAGACTTAAACATATGCTCA
 CCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTT
 15 TGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTT
 TTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGAC
 ATTTAAAGAAGACATTCAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTACATGAACATA
 TTGCAAATTTAGCTGGTAGCCCTGCTATTA AAAAAGGTATTTTACAGACTGTAAAAGTTGTT
 GATGAATTGGTCAAAGTAATGGGGCGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACG
 20 TGAAAATCAGACAACCTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAG
 AAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTG
 CAAAATGAAAAGCTCTATCTCTATTATCTCCAAAATGGAAGAGACATGTATGTGGACCAAGA
 ATTAGATATTAATCGTTTTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCCTTA
 AAGACGATTCAATAGACAATAAGGTCTTAACGCGTTCTGATAAAAAATCGTGGTAAATCGGAT
 25 AACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAACTATTGGAGACAACCTTCTAAACGC
 CAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTG
 AACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCAT
 GTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCG
 AGAGGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTCCGAAAAGATTTCCAAT
 30 TCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTC
 GTTGGAACGTGCTTTGATTAAGAAATATCCAAAACCTGAATCGGAGTTTGTCTATGGTGATTA
 TAAAGTTTATGATGTTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCG
 CAAAATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAAT
 GGAGAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGA

TAAAGGGCGAGATTTTGGCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCA
 AGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCG
 GACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAATATGGTGGTTTTGATAGTCC
 AACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAA
 5 AATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAATCCG
 ATTGACTTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAGTACC
 TAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAAT
 TACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAGT
 CATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAGCA
 10 GCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTT
 TAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAATA
 CGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGC
 TTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTGTAG
 ATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAG
 15 CTAGGAGGTGACTGA

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLEDSGETAEAT
 RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVD
 EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFI
 20 QLVQTYNQLFEEPNINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGL
 TPNFKNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPKEYKEIFFDQSKNGYAGYIDGGASQEEF
 YKFIKPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIHQIHLGELHAILRRQEDFYFPLK
 DNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
 25 NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKI IKDKDFLDNEENEDILEDIV
 LTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
 LKSDGFANRNFQMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
 30 QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNVIVKKTEVQTGGFSKESILPKRNS

DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSSFENP
IDFLEAKGYKEVKKDLIIKLPKYSLFELNGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLKDVL SAYNKH RD KPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDL SQ

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

In some embodiments, Cas9 refers to Cas9 from: *Corynebacterium ulcerans* (NCBI Refs: NC_015683.1, NC_017317.1); *Corynebacterium diphtheria* (NCBI Refs: NC_016782.1, NC_016786.1); *Spiroplasma syrphidicola* (NCBI Ref: NC_021284.1); *Prevotella intermedia* (NCBI Ref: NC_017861.1); *Spiroplasma taiwanense* (NCBI Ref: NC_021846.1); *Streptococcus iniae* (NCBI Ref: NC_021314.1); *Belliella baltica* (NCBI Ref: NC_018010.1); *Psychroflexus torquisI* (NCBI Ref: NC_018721.1); *Streptococcus thermophilus* (NCBI Ref: YP_820832.1), *Listeria innocua* (NCBI Ref: NP_472073.1), *Campylobacter jejuni* (NCBI Ref: YP_002344900.1) or *Neisseria meningitidis* (NCBI Ref: YP_002342100.1) or to a Cas9 from any other organism.

15 It should be appreciated that additional Cas9 proteins (*e.g.*, a nuclease dead Cas9 (dCas9), a Cas9 nickase (nCas9), or a nuclease active Cas9), including variants and homologs thereof, are within the scope of this disclosure. Exemplary Cas9 proteins include, without limitation, those provided below. In some embodiments, the Cas9 protein is a nuclease dead Cas9 (dCas9). In some embodiments, the Cas9 protein is a Cas9 nickase (nCas9). In some
20 embodiments, the Cas9 protein is a nuclease active Cas9.

In some embodiments, the Cas9 domain is a nuclease-inactive Cas9 domain (dCas9). For example, the dCas9 domain may bind to a duplexed nucleic acid molecule (*e.g.*, via a gRNA molecule) without cleaving either strand of the duplexed nucleic acid molecule. In some embodiments, the nuclease-inactive dCas9 domain comprises a D10X mutation and a
25 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. As one example, a nuclease-inactive Cas9
30 domain comprises the amino acid sequence set forth in Cloning vector pPlatTET-gRNA2 (Accession No. BAV54124).

The amino acid sequence of an exemplary catalytically inactive Cas9 (dCas9) is as follows:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
 RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFFHRLEESFLVEEDKKHERHPIFGNIVD
 EVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHAMIKFRGHFLIEGDLNPDNSDVKLFI
 QLVQTYNQLFEEPNINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALLSLGL
 5 TPNFKSNFDLAEDAQLQSKDQYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPKEYKEIFFDQSKNGYAGYIDGGASQEEF
 YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIHQIHLGELHAILRRQEDFYFPLK
 DNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMT
 NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
 10 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIV
 LTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
 LKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVV
 DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEKGELGSQILKEHPVENTQL
 QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSD
 15 NVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
 VAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
 VGTALIKKYPKLESEFVYGDKVYDVRKMIKSEQEIGKATKYFFYSNIMNFFKTEITLAN
 GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKRNS
 DKLIARKKDWDPKKGFFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSSFENP
 20 IDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
 HYEKLKGSPEQKQKLFVEQHKHYLDEIIIEQISEFSKRVILADANLDKVL SAYNKHDKPI
 REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSIITGLYETRIDLQ
 LGGD

(see, e.g., Qi *et al.*, “Repurposing CRISPR as an RNA-guided platform for sequence-specific
 25 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
 30 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, 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). A nuclease-inactivated Cas9 protein may interchangeably be referred to as a “dCas9” protein (for nuclease-“dead” Cas9) or catalytically inactive Cas9. Methods for generating a Cas9 protein (or a fragment thereof) having an inactive DNA cleavage domain are known (*See, e.g.*, Jinek *et al.*, *Science*. 337:816-821(2012); Qi *et al.*, “Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression” (2013) *Cell*. 28;152(5):1173-83, the entire contents of each of which are incorporated herein by reference). For example, the DNA cleavage domain of Cas9 is known to include two subdomains, the HNH nuclease subdomain and the RuvC1 subdomain. The HNH subdomain cleaves the strand complementary to the gRNA, whereas the RuvC1 subdomain cleaves the non-complementary strand. Mutations within these subdomains can silence the nuclease activity of Cas9. For example, the mutations D10A and H840A completely inactivate the nuclease activity of *S. pyogenes* Cas9 (Jinek *et al.*, *Science*. 337:816-821(2012); Qi *et al.*, *Cell*. 28;152(5):1173-83 (2013)).

In some embodiments, the dCas9 domain 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 dCas9 domains provided herein. In some embodiments, the Cas9 domain comprises an amino acid sequences that has 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, 50 or more or more mutations compared to any one of the amino acid sequences set forth herein. In some embodiments, the Cas9 domain comprises an amino acid sequence that has at least 10, at least 15, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 150, at least 200, at least 250, at least 300, at least 350, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, at least 1000, at least 1100, or at least 1200 identical contiguous amino acid residues as compared to any one of the amino acid sequences set forth herein.

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, the dCas9 comprises the amino acid sequence of dCas9 (D10A and H840A):

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
 RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVD
 10 EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFI
 QLVQTYNQLFENPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIALSLGL
 TPNFKSNFDLAEDAKLQLSKDQYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLEPKYKEIFFDQSKNGYAGYIDGGASQEEF
 YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIHQIHLGELHAILRRQEDFYFPLK
 15 DNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASQSFIERMT
 NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIV
 LTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
 LKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAICKGILQTVKVV
 20 DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETROITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEI GKATAKYFFYSNIMNFFKTEITLAN
 25 GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNS
 DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEKFN
 IDFLEAKGYKEVKKDLIIKLPKYSLFELNGRKRMLASAGELQKGNELALPSKYVNFLYLAS
 HYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFSKRVILADANLDKVL SAYNKHDKPI
 REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDL SQ
 30 LGGD (single underline: HNH domain; double underline: RuvC domain).

In some embodiments, the amino acid sequence of an exemplary catalytically inactive Cas9 (dCas9) is as follows:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
 RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVD

EVAYHEKYPTIYHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFI
 QLVQTYNQ LFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGL
 TPNFKNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQ LPEKYKEIFFDQSKNGYAGYIDGGASQEEF
 5 YKFIKPILEKMDGTEELLV KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLK
 DNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMT
 NFDKNLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLKIIKDKDFLDNEENEDILEDIV
 LTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
 10 LKSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA IKKGILQTVKVV
 DELVKVMGRHKPENIV IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
 QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSD
 NVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLT KAERGGLSELDKAGFIKRQLVETRQITKH
 VAQILDSRMNTKYDENDKLIREVKVITL KSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
 15 VGTALIKKYPKLESEFVYGDYKVYDVRKMI AKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
 GEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNS
 DKLIARKKDWDPKKYGGFDSPTVAYSVLV VAKVEKGKSKKLKSVKELLGITIMERSSSFENP
 IDFLEAKGYKEVKKDLIIKLPKYSLFEL ENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
 HYEKLKGS PEDNEQKQLFVEQH KHYLDEII EQISEFSKRVILADANLDKVL SAYNKH RDKPI
 20 REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ SITGLYETRIDL SQ
 LGGD

(see, e.g., Qi *et al.*, “Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.” *Cell*. 2013; 152(5):1173-83, the entire contents of which are incorporated herein by reference).

25 In some embodiments, the amino acid sequence of an exemplary catalytically inactive Cas9 (dCas9) is as follows:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHS IKKNLIGALLFDSGETAEAT
 RLKRTARRRYTRKKNRICYLQEIFS NEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVD
 EVAYHEKYPTIYHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFI
 30 QLVQTYNQ LFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGL
 TPNFKNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQ LPEKYKEIFFDQSKNGYAGYIDGGASQEEF
 YKFIKPILEKMDGTEELLV KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLK
 DNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMT

NFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
 LTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
 LKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVV
 5 DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
 QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSD
 NVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
 VAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
 VGTALIKKYPKLESEFVYGDKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
 10 GEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNS
 DKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSSFENP
 IDFLEAKGYKEVKKDLIIKLPKYSLFELNGRKRMLASAGELQKGNELALPSKYVNFLYLAS
 HYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFSKRVILADANLDKVL SAYNKH RDKPI
 REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDL SQ
 15 LGGD

In some embodiments, the Cas9 domain comprises a D10A mutation, while the residue at position 840 remains a histidine in the amino acid sequence provided above, or at corresponding positions in any of the amino acid sequences provided herein.

In other embodiments, dCas9 variants having mutations other than D10A and H840A
 20 are provided, which, *e.g.*, result in nuclease inactivated Cas9 (dCas9). Such mutations, by
 way of example, include other amino acid substitutions at D10 and H840, or other
 substitutions within the nuclease domains of Cas9 (*e.g.*, substitutions in the HNH nuclease
 subdomain and/or the RuvC1 subdomain). In some embodiments, variants or homologues of
 dCas9 are provided which are at least about 70% identical, at least about 80% identical, at
 25 least about 90% identical, at least about 95% identical, at least about 98% identical, at least
 about 99% identical, at least about 99.5% identical, or at least about 99.9% identical. In
 some embodiments, variants of dCas9 are provided having amino acid sequences which are
 shorter, or longer, by about 5 amino acids, by about 10 amino acids, by about 15 amino acids,
 by about 20 amino acids, by about 25 amino acids, by about 30 amino acids, by about 40
 30 amino acids, by about 50 amino acids, by about 75 amino acids, by about 100 amino acids or
 more.

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*
 5 reference).

In some embodiments, the Cas9 domain is a Cas9 nickase. 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
 10 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
 15 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
 20 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:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
 25 RLKRTARRRYTRKNNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPHIFGNIVD
 EVAYHEKYPTIYHLRKKLVDSTDKADLRLLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFI
 QLVQTYNQLFEEPNINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIALLSLGL
 TPNFKSNFDLAEDAKLQLSKDLYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLEPKYKEIFFDQSKNGYAGYIDGGASQEEF
 30 YKFIKPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPFLK
 DNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASQAQSFIERMT
 NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIV
 LTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF

LKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVV
 DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
 QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
 NVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
 5 VAQILDSRMNTKYDENDKLIREVKVITLKSILVSDFRKDFQFYKVREINNYHHAHDAYLNAV
 VGTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
 GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKRNS
 DKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEFKNP
 IDFLEAKGYKEVKKDLIIKLPKYSLFELNGRKRMLASAGELQKGNELALPSKYVNFLYLAS
 10 HYEKLKGSPEQKQKLFVEQHKHYLDEIIIEQISEFSKRVILADANLDKVL SAYNKH RDKPI
 REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDL SQ
 LGGD

In some embodiments, Cas9 refers to a Cas9 from archaea (*e.g.*, nanoarchaea), which
 constitute a domain and kingdom of single-celled prokaryotic microbes. In some
 15 embodiments, the programmable nucleotide binding protein may be a CasX or CasY protein,
 which have been described in, for example, Burstein *et al.*, "New CRISPR-Cas systems from
 uncultivated microbes." Cell Res. 2017 Feb 21. doi: 10.1038/cr.2017.21, the entire contents
 of which is hereby incorporated by reference. Using genome-resolved metagenomics, a
 number of CRISPR-Cas systems were identified, including the first reported Cas9 in the
 20 archaeal domain of life. This divergent Cas9 protein was found in little-studied nanoarchaea
 as part of an active CRISPR-Cas system. In bacteria, two previously unknown systems were
 discovered, CRISPR-CasX and CRISPR-CasY, which are among the most compact systems
 yet discovered. In some embodiments, in a base editor system described herein Cas9 is
 replaced by CasX, or a variant of CasX. In some embodiments, in a base editor system
 25 described herein Cas9 is replaced by CasY, or a variant of CasY. It should be appreciated that
 other RNA-guided DNA binding proteins may be used as a nucleic acid programmable DNA
 binding protein (napDNAbp), and are within the scope of this disclosure.

In some embodiments, the nucleic acid programmable DNA binding protein
 (napDNAbp) of any of the fusion proteins provided herein may be a CasX or CasY protein.
 30 In some embodiments, the napDNAbp is a CasX protein. In some embodiments, the
 napDNAbp is a CasY 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 CasX or CasY protein. In some embodiments, the

programmable nucleotide binding protein is a naturally-occurring CasX or CasY protein. In some embodiments, the programmable nucleotide binding protein 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 CasX or CasY protein described herein. It should be appreciated that CasX and CasY from other bacterial species may also be used in accordance with the present disclosure.

An exemplary CasX ((uniprot.org/uniprot/F0NN87; uniprot.org/uniprot/F0NH53) tr|F0NN87|F0NN87_SULIHCRISPR-associatedCasx protein OS = *Sulfolobus islandicus* (strain HVE10/4) GN = SiH_0402 PE=4 SV=1) amino acid sequence is as follows:

MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVNLNLAYKIAKNNEDAAAERRGK
AKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAP
SFVKPEFYEFGRSPGMVERTRRVKLEVEPHYLI IAAAGWVLTRLGKAKVSEG DYVGVNVFT P
TRGILYSLIQNVNGIVPGIKPETAFLWIARKVSSVTNPNVSVVRIYITISDAVGQNPTTIN
GGFSIDLTKLLEKRYLLSERLEAIARNALSISSNMRERYIVLANIYIEYLTG SKRLEDLLY
FANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG.

An exemplary CasX (>tr|F0NH53|F0NH53_SULIR CRISPR associated protein, Casx OS = *Sulfolobus islandicus* (strain REY15A) GN=SiRe_0771 PE=4 SV=1) amino acid sequence is as follows:

MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVNLNLAYKIAKNNEDAAAERRGK
AKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAP
SFVKPEFYKFGRSPGMVERTRRVKLEVEPHYLI IAAAGWVLTRLGKAKVSEG DYVGVNVFT P
TRGILYSLIQNVNGIVPGIKPETAFLWIARKVSSVTNPNVSVVSIYITISDAVGQNPTTIN
GGFSIDLTKLLEKRDLLSERLEAIARNALSISSNMRERYIVLANIYIEYLTGSKRLEDLLYF
ANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG.

Deltaproteobacteria CasX

MEKRINKIRKKLSADNATKPVSRSGPMKTLVVRVMTDDLKKRLEKRRKKPEVMPQVISNNAA
NNLRMLLDDYTKMKEAILQVYWQEFKDDHVGLMCKFAQPASKKIDQNKLKPEMDEKGNLTTA
GFACSQCGQPLFVYKLEQVSEK GKAYTNYFGRCNVAEHEKLILLQLKPKVSD EAVTYSLG
KFGQRALDFYSIHVTKESTHPVKPLAQIAGNRYASGPVGKALSDACMGTIASFLSKYQDII I
EHQKVVKGNQKRLESRELAKENLEYPSTLPPQPHTKEGVDFAYNEVIARVRMWNLNLNW
QKLKLSRDDAKPLRLKGFPSFPVVERRENEVDWNTINEVKKLIDAKRDMGRVFWSGVTAE
KRNTILEGYNILPNENDHKKREGSLENPKKPAKRQFGDLLLYLEKKYAGDWGKVFDEAWERI
DKKIAGLTSHIEREEARNAEDAQSKAVLTDWLRKASFVLERLKEMDEKEFYACEIQLOKWY

GDLRGNPF AVEAENRVVDISGFSIGSDGHSIQYRNLLAWKYLENGKREFYLLMNYGKKGRIR
 FTDGTDIKKSGKWQGLLYGGGKAKVIDLTFDPDDEQLIILPLAFGTRQGREFIWNLLSLET
 GLIKLANGRVIEKTIYNKKIGRDEPALFVALTFERREVDPSPNIKPVNLIGVARGENIPAVI
 ALTDPGECPLPEFKDSSGGPTDILRIGEGYKEKQRAIQAAKEVEQRRAGGYSRKFASKSRNL
 5 ADDMVRNSARDLFYHAVTHDAVLVFANLSRGFGRQGKRTFMTERQYTKMEDWLTAKLAYEGL
 TSKTYLSKTLAQYTSKTC SNCGFTITYADMDVMLVRLKKTSDGWATTNNKELKA EYQITYY
 NRYKRQTV EKELSAELDRLSEESGNNDISKWTKGRRDEALFLLKKRFSHRPVQE QFVCLDCG
 HEVHAAEQ AALNIARSWLFLNSNSTEFKSYKSGKQPFVGAWQAFYKRRLKEVWKPN A

An exemplary CasY ((ncbi.nlm.nih.gov/protein/APG80656.1) >APG80656.1

10 CRISPR-associated protein CasY [uncultured Parcubacteria group bacterium]) amino acid
 sequence is as follows:

MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIVSAINDDYVGL
 YGLSNFDDLYNAEKRNEEKVYSVLDWFYDCVQYGAVFSYTAPGLLKNAEVRGGSYELTKTL
 KGSHLYDELQIDKVIKFLNKKEISRANGSLDKLKKDIIDCFKA EYRERHKDQCNKLADDIKN
 15 AKKDAGASLGERQKKLFRDFFGISEQSENDKPSFTNPLNLTCCLLPFDTVNNNRNRGEVLFN
 KLKEYAQKLDKNEGSLEMWEYIGIGNSGTAFSNFLGEGFLGRLRENKITE LKKAMMDITDAW
 RGQEQEEELEKRLRILAALTIKLREP KFDNHWGGYRS DINGKLSSWLQNYINQTVKIKEDLK
 GHKKDLKKAKEMINRFGESDTKEEAVVSSLLESIEKIVPDD SADDEKPDIPAIAIYRRFLSD
 GR LTLNRFVQREDVQEALIKERLEAEK KKKPKKRKKKS DAEDEKETIDFKELFPHLAKPLKL
 20 VPNFYGDSKRELYKKYKNAAIYTDALWKAVEKIYKSAFSSSLKNSFFD TDFDKDFFIKRLQK
 IFSVYRRFNTDKWKPIVKN SFAPYCDIVSLAENEVLYKPKQSR SRKSA AIDKNRVRLPSTEN
 IAKAGIALARELSVAGFDWKDLLKKEEHEEYIDLIELHKTALALLAVTETQLDISALDFVE
 NGTVKDFMKT RDGNLVLEGRFLEMFSQSIVFSELRGLAGLMSRKEFITRS AIQTMNGKQAE L
 LYIPHEFQSAKITTPKEMSRAFLDLAPAEFATSLEPESLSEKSL LKLKQMRYPHYFGYELT
 25 RTGQGIDGGVAENALRLEKSPVKKREIKCKQYKTLGRGQNKIVLYVRSSYYQTQFLEWFLHR
 PKNVQTDVAVSGSFLIDEKKVKTRWNYDALTVALEPVSGSERVFVSQPFTIFPEKSAEEEGQ
 RYL GIDIGEYGIAYTALEITGDSAKILDQNFISDPQLKTLREEVKGLKLDQRRGTFAMPSTK
 IARIRESLVHSLRNRIHHLALKHKAKIVYELEVS RFEEGKQKIKKVYATLKKADVYSEIDAD
 KNLQTTVWGKLAVASEISASYTSQFCGACKKLWRAEMQVDETITTQELIGTVRVIKGGTLID
 30 AIKDFMRPPIFDENDTPFPKYRDFCDKHHISKMRGNSCLFICPF CRANADADIQASQTIAL
 LRYVKEEKKV EDYFERFRKLKNIKVLGQM KKI.

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.

5 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 cases, 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
10 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
15 and “b” and “c” are the cleavage products).

 In some cases, 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
20 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.
25 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
30 mutations. In most cases, 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
5 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
10 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
15 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
20 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.

In some cases, Cas9 is a variant Cas9 protein. A variant Cas9 polypeptide has an
25 amino acid sequence that is different by one amino acid (*e.g.*, has a deletion, insertion, substitution, fusion) when compared to the amino acid sequence of a wild type Cas9 protein. In some instances, the variant Cas9 polypeptide has an amino acid change (*e.g.*, deletion, insertion, or substitution) that reduces the nuclease activity of the Cas9 polypeptide. For example, in some instances, the variant Cas9 polypeptide has less than 50%, less than 40%,
30 less than 30%, less than 20%, less than 10%, less than 5%, or less than 1% of the nuclease activity of the corresponding wild-type Cas9 protein. In some cases, the variant Cas9 protein has no substantial nuclease activity. When a subject Cas9 protein is a variant Cas9 protein that has no substantial nuclease activity, it can be referred to as “dCas9.”

In some cases, a variant Cas9 protein has reduced nuclease activity. For example, a variant Cas9 protein exhibits less than about 20%, less than about 15%, less than about 10%, less than about 5%, less than about 1%, or less than about 0.1%, of the endonuclease activity of a wild-type Cas9 protein, *e.g.*, a wild-type Cas9 protein.

5 In some cases, 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
10 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;
15 337(6096):816-21).

In some cases, 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
20 (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
25 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).

In some cases, a variant Cas9 protein has a reduced ability to cleave both the complementary and the non-complementary strands of a double stranded target DNA. As a
30 non-limiting example, in some cases, the variant Cas9 protein harbors both the D10A and the H840A mutations such that the polypeptide has a reduced ability to cleave both the complementary and the non-complementary strands of a double stranded 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 cases, 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 cases, 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 cases, 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 cases, 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 cases, 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 cases, 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 cases, 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 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 cases, when such a variant Cas9 protein is used in a method of binding, the method can
 5 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,
 10 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
 15 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-MQKFRAER, spCas9-MQKSER, spCas9-LRKIQK, or spCas9-LRVSQL.

20 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' is 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
 25 *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
 30 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 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 more similar to types I and III than from type II systems. Functional Cpf1 doesn't need 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 (proximately 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' in contrast to the G-rich PAM targeted by Cas9. After identification of PAM, Cpf1 introduces a sticky-end-like DNA double-stranded break of 4 or 5 nucleotides overhang.

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 variant 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

5 corresponding position thereof. Exemplary amino acid substitutions and PAM specificity of SpCas9 variants are shown in **Tables 1A-1D**.

Table 1A

SpCas9	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

10

Table 1B

SpCas9	SpCas9 amino acid position																		
	11	11	11	11	11	11	11	11	12	12	12	12	12	12	13	13	13	13	13
	14	34	35	37	39	51	80	88	11	19	21	56	64	90	18	17	20	23	33
	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 1C

SpCas9	SpCas9 amino acid position																			
	11 14	11 31	11 35	11 50	11 56	11 80	11 91	12 18	12 19	12 21	12 27	12 49	12 53	12 86	12 93	13 20	13 21	13 32	13 35	13 39
	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		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 1D

SpCas9	SpCas9 amino acid position													
	111	112	113	118	120	121	123	128	130	133	133	133	133	134
	4	7	5	0	7	9	4	6	1	2	5	7	8	9
	R	D	D	D	E	E	N	N	P	D	R	T	S	H
SacB.CA C			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		
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

In some embodiments, the Cas9 is a *Neisseria meningitidis* Cas9 (NmeCas9) or a variant thereof. In some embodiments, the NmeCas9 has specificity for a NNNNGAYW PAM, wherein Y is C or T and W is A or T. In some embodiments, the NmeCas9 has specificity for a NNNNGYTT PAM, wherein Y is C or T. In some embodiments, the NmeCas9 has specificity for a NNNNGTCT PAM. In some embodiments, the NmeCas9 is a Nme1 Cas9. In some embodiments, the NmeCas9 has specificity for a NNNNGATT PAM, a NNNNCCTA PAM, a NNNNCCTC PAM, a NNNNCCTT PAM, a NNNNCCTG PAM, a NNNNCCGT PAM, a NNNNCCGGPAM, a NNNNCCCA PAM, a NNNNCCCT PAM, a NNNNCCCC PAM, a NNNNCCAT PAM, a NNNNCCAG PAM, a NNNNCCAT PAM, or a NNNGATT PAM. In some embodiments, the Nme1Cas9 has specificity for a NNNNGATT PAM, a NNNNCCTA PAM, a NNNNCCTC PAM, a NNNNCCTT PAM, or a NNNNCCTG PAM. In some embodiments, the NmeCas9 has specificity for a CAA PAM, a CAAA PAM, or a CCA PAM. In some embodiments, the NmeCas9 is a Nme2 Cas9. In some embodiments, the NmeCas9 has specificity for a NNNNCC (N4CC) PAM, wherein N is any one of A, G, C, or T. In some embodiments, the NmeCas9 has specificity for a NNNNCCGT PAM, a NNNNCCGGPAM, a NNNNCCCA PAM, a NNNNCCCT PAM, a NNNNCCCC PAM, a NNNNCCAT PAM, a NNNNCCAG PAM, a NNNNCCAT PAM, or a NNNGATT PAM. In some embodiments, the NmeCas9 is a Nme3Cas9. In some embodiments, the NmeCas9 has specificity for a NNNNCAAA PAM, a NNNNCC PAM, or a NNNNCNNN

PAM. Additional NmeCas9 features and PAM sequences as described in Edraki *et al.* Mol. Cell. (2019) 73(4): 714-726 is incorporated herein by reference in its entirety.

An exemplary amino acid sequence of a Nme1Cas9 is provided below:

type II CRISPR RNA-guided endonuclease Cas9 [*Neisseria meningitidis*] WP_002235162.1

```

5      1 maafknpin yilgldigia svgwamveid edenpiclid lgvrvferae vpktgdslam
      61 arrlarsvrr ltrrrahrll rarrllkreg vlqaadfdn glikslpntp wqlraaaldr
      121 kltplewsav llhlikhrgy lsqrkneget adkelgallk gvadnahalq tgdfrtpael
      181 alnkfekesg hirnqrgdys htfsrkdlqa elillfekqk efgnphvsgg lkegietllm
      241 tqrpalsgda vqkmlghctf epaepkaakn tytaerfiwl tklnnlrile ggserpltdt
10    301 eratlmdepy rkskltyaqa rklglgledta ffgkglrygkd naeastlmem kayhaisral
      361 ekeglkdkks plnlspelqd eigtafslfk tdeditgrlk driqpeilea llkhisfdkf
      421 vqislkalrr ivplmeqgkr ydeacaeiyg dhygkntee kiylppipad eirnpvvlra
      481 lsqarkving vrrygspar ihietarevg ksfkdrkeie krqeenrkdr ekaaakfrey
      541 fpnfvgpks kdilklrlye qqhgkclysg keinlgrlne kgyveidhal pfsrtwddsf
15    601 nnkvvlvgse nqnkgnqtpy eyfngkdnsr ewqefkarve tsrfprskkq rillqkfdded
      661 gfkernlndt ryvnrflcqf vadrmlrtgk gkkrvfasng qitnllrgfw glrkvraend
      721 rhhaldavvv acstvamqqk itrfrvykem nafdgktidk etgevlhqkt hfpqpweffa
      781 qevmirvfgk pdgkpefeea dtpeklrtll aeklssrpea vheyvtplfv srpnrkmsg
      841 qghmetvksa krldegsvl rvpltqlklk dlekmvnrer epklyealka rleahkddpa
20    901 kafaepfyky dkagnrtqqv kavreveqvqk tgvwvrnhng iadnatmvrv dvfekgdkyy
      961 lvpiyswqva kgilpdravv qgkdeedwql iddsfnfkfs lhpndlvevi tkkarmfgyf
      1021 aschrgtgni nirihldhk igkngilegi gvktalsfqk yqidelgkei rpcrlkkrpp
      1081 vr

```

An exemplary amino acid sequence of a Nme2Cas9 is provided below:

25 type II CRISPR RNA-guided endonuclease Cas9 [*Neisseria meningitidis*] WP_002230835.1

```

      1 maafknpin yilgldigia svgwamveid eenpirlid lgvrvferae vpktgdslam
      61 arrlarsvrr ltrrrahrll rarrllkreg vlqaadfdn glikslpntp wqlraaaldr
      121 kltplewsav llhlikhrgy lsqrkneget adkelgallk gvannahalq tgdfrtpael
      181 alnkfekesg hirnqrgdys htfsrkdlqa elillfekqk efgnphvsgg lkegietllm
30    241 tqrpalsgda vqkmlghctf epaepkaakn tytaerfiwl tklnnlrile ggserpltdt
      301 eratlmdepy rkskltyaqa rklglgledta ffgkglrygkd naeastlmem kayhaisral
      361 ekeglkdkks plnlsselqd eigtafslfk tdeditgrlk drvqpeilea llkhisfdkf
      421 vqislkalrr ivplmeqgkr ydeacaeiyg dhygkntee kiylppipad eirnpvvlra
      481 lsqarkving vrrygspar ihietarevg ksfkdrkeie krqeenrkdr ekaaakfrey
35    541 fpnfvgpks kdilklrlye qqhgkclysg keinlvrlne kgyveidhal pfsrtwddsf
      601 nnkvvlvgse nqnkgnqtpy eyfngkdnsr ewqefkarve tsrfprskkq rillqkfdded
      661 gfkecnlndt ryvnrflcqf vadhilltgk gkrrvfasng qitnllrgfw glrkvraend
      721 rhhaldavvv acstvamqqk itrfrvykem nafdgktidk etgkvlhqkt hfpqpweffa
      781 qevmirvfgk pdgkpefeea dtpeklrtll aeklssrpea vheyvtplfv srpnrkmsg

```


841 ahkdtlrsak rfvkhnekis vkrvwlteik ladlenmvny kngreielye alkarleayg
 901 gnakqafdpk dnpfykkggq lvkavrvekt qesgvllnkk naytiadngd mrvvdvfckv
 961 dkkgnqyfi vpiyawqvae nilpdidckg yriddsytfc fslhkydlia fqkdekskve
 1021 fayyincdss ngrfylawhd kgskeqqfri stqnlvliqk yqvnelgkei rpcrlkkrpp
 5 1081 vr

Cas12 domains of Nucleobase Editors

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, albeit different types (Type II and Type V, respectively). In addition to Cpf1, Class 2, Type V CRISPR-Cas systems also comprise Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i and Cas12j/CasΦ). See, *e.g.*, Shmakov *et al.*, “Discovery and Functional Characterization of Diverse Class 2 CRISPR Cas Systems,” Mol. Cell, 2015 Nov. 5; 60(3): 385-397; Makarova *et al.*, “Classification and Nomenclature of CRISPR-Cas Systems: Where from Here?” CRISPR Journal, 2018, 1(5): 325-336; and 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. Type V Cas proteins contain a RuvC (or RuvC-like) endonuclease domain. While production of mature CRISPR RNA (crRNA) is generally tracrRNA-independent, Cas12b/C2c1, for example, requires tracrRNA for production of crRNA. Cas12b/C2c1 depends on both crRNA and tracrRNA for DNA cleavage.

Nucleic acid programmable DNA binding proteins contemplated in the present invention include Cas proteins that are classified as Class 2, Type V (Cas12 proteins). Non-limiting examples of Cas Class 2, Type V proteins include Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, and Cas12j/CasΦ homologues thereof, or modified versions thereof. As used herein, a Cas12 protein can also be referred to as a Cas12 nuclease, a Cas12 domain, or a Cas12 protein domain. In some embodiments, the Cas12 proteins of the present invention comprise an amino acid sequence interrupted by an internally fused protein domain such as a deaminase domain.

In some embodiments, the Cas12 domain is a nuclease inactive Cas12 domain or a Cas12 nickase. In some embodiments, the Cas12 domain is a nuclease active domain. For example, the Cas12 domain may be a Cas12 domain that nicks one strand of a duplexed nucleic acid (*e.g.*, duplexed DNA molecule). In some embodiments, the Cas12 domain

comprises any one of the amino acid sequences as set forth herein. In some embodiments the Cas12 domain 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 herein. In some embodiments, the Cas12 domain 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, 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 mutations compared to any one of the amino acid sequences set forth herein. In some embodiments, the Cas12 domain comprises an amino acid sequence that has at least 10, at least 15, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 150, at least 200, at least 250, at least 300, at least 350, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, at least 1000, at least 1100, or at least 1200 identical contiguous amino acid residues as compared to any one of the amino acid sequences set forth herein.

In some embodiments, proteins comprising fragments of Cas12 are provided. For example, in some embodiments, a protein comprises one of two Cas12 domains: (1) the gRNA binding domain of Cas12; or (2) the DNA cleavage domain of Cas12. In some embodiments, proteins comprising Cas12 or fragments thereof are referred to as “Cas12 variants.” A Cas12 variant shares homology to Cas12, or a fragment thereof. For example, a Cas12 variant is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to wild type Cas12. In some embodiments, the Cas12 variant may have 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, 50 or more amino acid changes compared to wild type Cas12. In some embodiments, the Cas12 variant comprises a fragment of Cas12 (*e.g.*, a gRNA binding domain or a DNA cleavage domain), such that the fragment is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to the corresponding fragment of wild type Cas12. In some embodiments, the fragment is 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% identical, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% of the amino acid length of a corresponding wild type Cas12. In some embodiments, the fragment is at least 100 amino acids in length. In some embodiments, the fragment is at least 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, or at least 1300 amino acids in length.

In some embodiments, Cas12 corresponds to, or comprises in part or in whole, a Cas12 amino acid sequence having one or more mutations that alter the Cas12 nuclease activity. Such mutations, by way of example, include amino acid substitutions within the RuvC nuclease domain of Cas12. In some embodiments, variants or homologues of Cas12 are provided which are at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to a wild type Cas12. In some embodiments, variants of Cas12 are provided having amino acid sequences which are shorter, or longer, by about 5 amino acids, by about 10 amino acids, by about 15 amino acids, by about 20 amino acids, by about 25 amino acids, by about 30 amino acids, by about 40 amino acids, by about 50 amino acids, by about 75 amino acids, by about 100 amino acids or more.

In some embodiments, Cas12 fusion proteins as provided herein comprise the full-length amino acid sequence of a Cas12 protein, *e.g.*, one of the Cas12 sequences provided herein. In other embodiments, however, fusion proteins as provided herein do not comprise a full-length Cas12 sequence, but only one or more fragments thereof. Exemplary amino acid sequences of suitable Cas12 domains are provided herein, and additional suitable sequences of Cas12 domains and fragments will be apparent to those of skill in the art.

Generally, the class 2, Type V Cas proteins have a single functional RuvC endonuclease domain (See, *e.g.*, Chen *et al.*, “CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity,” *Science* 360:436-439 (2018)). In some cases, the Cas12 protein is a variant Cas12b protein. (See Strecker *et al.*, *Nature Communications*, 2019, 10(1): Art. No.: 212). In one embodiment, a variant Cas12 polypeptide has an amino acid sequence that is different by 1, 2, 3, 4, 5 or more amino acids (*e.g.*, has a deletion, insertion, substitution, fusion) when compared to the amino acid sequence of a wild type Cas12 protein. In some instances, the variant Cas12 polypeptide has an amino acid change (*e.g.*, deletion, insertion, or substitution) that reduces the activity of the Cas12 polypeptide.

For example, in some instances, the variant Cas12 is a Cas12b polypeptide that has less than 50%, less than 40%, less than 30%, less than 20%, less than 10%, less than 5%, or less than 1% of the nickase activity of the corresponding wild-type Cas12b protein. In some cases, the variant Cas12b protein has no substantial nickase activity.

5 In some cases, a variant Cas12b protein has reduced nickase activity. For example, a variant Cas12b protein exhibits less than about 20%, less than about 15%, less than about 10%, less than about 5%, less than about 1%, or less than about 0.1%, of the nickase activity of a wild-type Cas12b protein.

In some embodiments, the Cas12 protein includes RNA-guided endonucleases from the Cas12a/Cpf1 family that displays 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 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'-TTTN-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 aspects of the present invention, a vector 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. Cas12 can refer to a polypeptide 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 Cas12 polypeptide (e.g., Cas12 from *Bacillus hisashii*). Cas12 can refer to a polypeptide with at most or at most 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 Cas12 polypeptide (e.g., from *Bacillus hisashii* (BhCas12b), *Bacillus* sp. V3-13 (BvCas12b), and *Alicyclobacillus acidiphilus* (AaCas12b)). Cas12 can refer to the wild type or a modified form of the Cas12 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, BhCas12b guide polynucleotide has the following sequence:

BhCas12b sgRNA scaffold (underlined) + 20nt to 23nt guide sequence (denoted by N_n)

5' GUUCUGTCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGCAGGGUGUGAGAAACUC
CUAUUGCUGGACGAUGUCUCUUACGAGGCAUAGCACNNNNNNNNNNNNNNNNNNNN-3'

In some embodiments, BvCas12b and AaCas12b guide polynucleotides have the following sequences:

BvCas12b sgRNA scaffold (underlined) + 20nt to 23nt guide sequence (denoted by N_n)

5' GACCUAUAGGGUCAUGAAUCUGUGCGUGUGCCAUAGUAAUUAUAAAAUUACCCACCACA
GGAGCACCUGAAAACAGGUGCUUUGGCACNNNNNNNNNNNNNNNNNNNN-3'

AaCas12b sgRNA scaffold (underlined) + 20nt to 23nt guide sequence (denoted by N_n)

5' GUCUAAAGGACAGAAUUUUUCAACGGGUGUGCCAAUGGCCACUUUCCAGGUGGCCAAAGCC
CGUUGAACUUCUCAAAAAGAACGAUCUGAGAAGUGGCACNNNNNNNNNNNNNNNNNNNN-3'

Nucleic acid programmable DNA binding proteins

Some aspects of the disclosure provide fusion proteins comprising domains that act as nucleic acid programmable DNA binding proteins, which may be used to guide a protein, such as a base editor, to a specific nucleic acid (e.g., DNA or RNA) sequence. In particular embodiments, a fusion protein comprises a nucleic acid programmable DNA binding protein

domain and a deaminase domain. Non-limiting examples of nucleic acid programmable DNA binding proteins include, Cas9 (*e.g.*, dCas9 and nCas9), Cas12a/Cpf1, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i and Cas12j/CasΦ. 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/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, Cas12j/CasΦ, 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.

One example of a nucleic acid programmable DNA-binding protein that has different PAM specificity than Cas9 is Clustered Regularly Interspaced Short Palindromic Repeats from *Prevotella* and *Francisella* 1 (Cpf1). Similar to Cas9, Cpf1 is also a class 2 CRISPR effector. It has been shown that Cpf1 mediates robust DNA interference with features distinct from Cas9. Cpf1 is a single RNA-guided endonuclease lacking tracrRNA, and it utilizes a T-rich protospacer-adjacent motif (TTN, TTTN, or YTN). Moreover, Cpf1 cleaves DNA via a staggered DNA double-stranded break. Out of 16 Cpf1-family proteins, two enzymes from *Acidaminococcus* and *Lachnospiraceae* are shown to have efficient genome-editing activity in human cells. Cpf1 proteins are known in the art and have been described previously, for example Yamano *et al.*, “Crystal structure of Cpf1 in complex with guide RNA and target DNA.” *Cell* (165) 2016, p. 949-962; the entire contents of which is hereby incorporated by reference.

Useful in the present compositions and methods are nuclease-inactive Cpf1 (dCpf1) variants that may be used as a guide nucleotide sequence-programmable DNA-binding

protein domain. The Cpf1 protein has a RuvC-like endonuclease domain that is similar to the RuvC domain of Cas9 but does not have a HNH endonuclease domain, and the N-terminal of Cpf1 does not have the alfa-helical recognition lobe of Cas9. It was shown in Zetsche *et al.*, *Cell*, 163, 759-771, 2015 (which is incorporated herein by reference) that, the RuvC-like domain of Cpf1 is responsible for cleaving both DNA strands and inactivation of the RuvC-like domain inactivates Cpf1 nuclease activity. For example, mutations corresponding to D917A, E1006A, or D1255A in *Francisella novicida* Cpf1 inactivate Cpf1 nuclease activity. In some embodiments, the dCpf1 of the present disclosure comprises mutations corresponding to D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A, or D917A/E1006A/D1255A. It is to be understood that any mutations, *e.g.*, substitution mutations, deletions, or insertions that inactivate the RuvC domain of Cpf1, may be used in accordance with the present disclosure.

In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) of any of the fusion proteins provided herein may be a Cpf1 protein. In some embodiments, the Cpf1 protein is a Cpf1 nickase (nCpf1). In some embodiments, the Cpf1 protein is a nuclease inactive Cpf1 (dCpf1). In some embodiments, the Cpf1, the nCpf1, or the dCpf1 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 Cpf1 sequence disclosed herein. In some embodiments, the dCpf1 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 Cpf1 sequence disclosed herein, and comprises mutations corresponding to D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A, or D917A/E1006A/D1255A. It should be appreciated that Cpf1 from other bacterial species may also be used in accordance with the present disclosure.

Wild-type *Francisella novicida* Cpf1 (D917, E1006, and D1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLIIDDEKRAKDYKKAKQIIDKYHQFFI
 EEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFN
 QNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENR
 KNVYSSNDIPTSIIRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDY
 KTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQ
 INDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTMQSFYEQIAAFKTVEEKSIKE
 TLLSLFDDLKAQKLDLSKIYFKNDKSLTDLQQVFDDYSVIGTAVLEYITQQIAPKNLDNPS

KKEQELIAKKTEKAKYLSLETIKLAL EEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
 DNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNLLHKLKIFHISQSEDKANILDKD
 EHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAI
 LFIKDDKY YLGVMNKKNNKIFDDKA IENKGE GYKKIVYKLLPGANKMLPKVFFSAKSIKFY
 5 NPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDT
 QRYNSIDEFYREVENQGYKLT FENISESYIDS VVNQGKLYLFQIYNKDFSAYS SKGRPNLHTL
 YWKALFDERNLQDVVYKLNGEAELFYRKQSI PKKITHPAKEAIANKNDNPKKESVFEYDLI
 KDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLEKANDVHILSI DRGERHLAYYTLDV
 GKGNI IKQDTFNI IGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEI
 10 AKLVIEYNAIVVF EDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNLYLVFKDNEFDKTGGVLRA
 YQLTAPFETFKKMGKQTGI IYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKIC
 YNLDKGYFEFSFDYKNFGDKAAKGWTIASFGSRLINFRNSDKNHNWDTREVPYPTKELEKLL
 KDYSIEYGHGECIKAAICGESDKKFFAKLT SVLNTILQMRNSKTGTELDYLISPVADVNGNF
 FDSRQAPKNMPQDA DANGAYHIGLGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

15

Francisella novicida CpfI D917A (A917, E1006, and D1255 are bolded and underlined)

MSIQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
 EEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFN
 QNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEI IKSFKGWTTYFKGFHENR
 20 KNVYSSNDIPTSI IYRIVDDNL PKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDY
 KTSEVNQRVFSLDEVFEIANFN NYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQ
 INDKTLKKYKMSVLFKQILSDTESKS FVIDKLEDDSDVVTMQSFYEQIAAFKTVEEKSIKE
 TLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPS
 KKEQELIAKKTEKAKYLSLETIKLAL EEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
 25 DNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNLLHKLKIFHISQSEDKANILDKD
 EHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAI
 LFIKDDKY YLGVMNKKNNKIFDDKA IENKGE GYKKIVYKLLPGANKMLPKVFFSAKSIKFY
 NPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDT
 QRYNSIDEFYREVENQGYKLT FENISESYIDS VVNQGKLYLFQIYNKDFSAYS SKGRPNLHTL
 30 YWKALFDERNLQDVVYKLNGEAELFYRKQSI PKKITHPAKEAIANKNDNPKKESVFEYDLI
 KDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLEKANDVHILSI ARGERHLAYYTLDV
 GKGNI IKQDTFNI IGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEI
 AKLVIEYNAIVVF EDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNLYLVFKDNEFDKTGGVLRA
 YQLTAPFETFKKMGKQTGI IYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKIC

YNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLL
 KDYSIEYGHGECIKAAICGESDKKFFAKLTSLVNTILQMRNSKTGTELDYLISPADVNGNF
 FDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

5 *Francisella novicida* CpfI E1006A (D917, A1006, and D1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
 EEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFN
 QNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEI IKSFKGWTTYFKGFHENR
 KNVYSSNDIPTSI IYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDY
 10 KTSEVNQRVFSLDEVFEIANFN NYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQ
 INDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTMQSFYEQIAAFKTVEEKSIKE
 TLLSLFDDLKAQKLDLSKIYFKNDKSLTDLQQVFDDYSVIGTAVLEYITQQIAPKNLDNPS
 KKEQELIAKKTEKAKYLSLETIKLAL EEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
 DNLAQISIKYQNQGKKDLLQASAEDDVKA IKDLLDQTNNLLHKLKIFHISQSEDKANILDKD
 15 EHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAI
 LFIKDDKYLLGVMNKKNNKIFDDKA IKENKGEYKKIVYKLLPGANKMLPKVFFSAKSIKFY
 NPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDT
 QRYNSIDEFYREVENQGYKLT FENISESYIDSVVNQGLYLFQIYNKDFSAYSKGRPNLHTL
 YWKALFDERNLQDVVYKLNGEAE LFYRKQSI PKKITHPAKEAIANKNDNPKKESVFEYDLI
 20 KDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI DRGERHLAYYTIVD
 GKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEI
 AKLVIEYNAIVVFADLNFGFKRGRFKVEKQVYQKLEKMLIEKLNLYLVFKDNEFDKTGGVLRA
 YQLTAPFETFKKMGKQTGI IYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKIC
 YNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLL
 25 KDYSIEYGHGECIKAAICGESDKKFFAKLTSLVNTILQMRNSKTGTELDYLISPADVNGNF
 FDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

Francisella novicida CpfI D1255A (D917, E1006, and A1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
 30 EEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFN
 QNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEI IKSFKGWTTYFKGFHENR
 KNVYSSNDIPTSI IYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDY
 KTSEVNQRVFSLDEVFEIANFN NYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQ
 INDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTMQSFYEQIAAFKTVEEKSIKE

T L S L L F D D L K A Q K L D L S K I Y F K N D K S L T D L S Q Q V F D D Y S V I G T A V L E Y I T Q Q I A P K N L D N P S
 K K E Q E L I A K K T E K A K Y L S L E T I K L A L E E F N K H R D I D K Q C R F E E I L A N F A A I P M I F D E I A Q N K
 D N L A Q I S I K Y Q N Q G K K D L L Q A S A E D D V K A I K D L L D Q T N N L L H K L K I F H I S Q S E D K A N I L D K D
 E H F Y L V F E E C Y F E L A N I V P L Y N K I R N Y I T Q K P Y S D E K F K L N F E N S T L A N G W D K N K E P D N T A I
 5 L F I K D D K Y Y L G V M N K K N N K I F D D K A I K E N K G E G Y K K I V Y K L L P G A N K M L P K V F F S A K S I K F Y
 N P S E D I L R I R N H S T H T K N G S P Q K G Y E K F E F N I E D C R K F I D F Y K Q S I S K H P E W K D F G F R F S D T
 Q R Y N S I D E F Y R E V E N Q G Y K L T F E N I S E S Y I D S V V N Q G K L Y L F Q I Y N K D F S A Y S K G R P N L H T L
 Y W K A L F D E R N L Q D V V Y K L N G E A E L F Y R K Q S I P K K I T H P A K E A I A N K N K D N P K K E S V F E Y D L I
 K D K R F T E D K F F F H C P I T I N F K S S G A N K F N D E I N L L L K E K A N D V H I L S I D R G E R H L A Y Y T L V D
 10 G K G N I I K Q D T F N I I G N D R M K T N Y H D K L A A I E K D R D S A R K D W K K I N N I K E M K E G Y L S Q V V H E I
 A K L V I E Y N A I V V F E D L N F G F K R G R F K V E K Q V Y Q K L E K M L I E K L N Y L V F K D N E F D K T G G V L R A
 Y Q L T A P F E T F K K M G K Q T G I I Y Y V P A G F T S K I C P V T G F V N Q L Y P K Y E S V S K S Q E F F S K F D K I C
 Y N L D K G Y F E F S F D Y K N F G D K A A K G W T I A S F G S R L I N F R N S D K N H N W D T R E V Y P T K E L E K L L
 K D Y S I E Y G H G E C I K A A I C G E S D K K F F A K L T S V L N T I L Q M R N S K T G T E L D Y L I S P V A D V N G N F
 15 F D S R Q A P K N M P Q D A A A N G A Y H I G L K G L M L L G R I K N N Q E G K K L N L V I K N E E Y F E F V Q N R N N

Francisella novicida Cpfl D917A/E1006A (A917, A1006, and D1255 are bolded and underlined)

M S I Y Q E F V N K Y S L S K T L R F E L I P Q G K T L E N I K A R G L I L D D E K R A K D Y K K A K Q I I D K Y H Q F F I
 20 E E I L S S V C I S E D L L Q N Y S D V Y F K L K S D D D N L Q K D F K S A K D T I K K Q I S E Y I K D S E K F K N L F N
 Q N L I D A K K G Q E S D L I L W L K Q S K D N G I E L F K A N S D I T D I D E A L E I I K S F K G W T T Y F K G F H E N R
 K N V Y S S N D I P T S I I Y R I V D D N L P K F L E N K A K Y E S L K D K A P E A I N Y E Q I K K D L A E E L T F D I D Y
 K T S E V N Q R V F S L D E V F E I A N F N N Y L N Q S G I T K F N T I I G G K F V N G E N T K R K G I N E Y I N L Y S Q Q
 I N D K T L K K Y K M S V L F K Q I L S D T E S K S F V I D K L E D D S D V V T T M Q S F Y E Q I A A F K T V E E K S I K E
 25 T L S L L F D D L K A Q K L D L S K I Y F K N D K S L T D L S Q Q V F D D Y S V I G T A V L E Y I T Q Q I A P K N L D N P S
 K K E Q E L I A K K T E K A K Y L S L E T I K L A L E E F N K H R D I D K Q C R F E E I L A N F A A I P M I F D E I A Q N K
 D N L A Q I S I K Y Q N Q G K K D L L Q A S A E D D V K A I K D L L D Q T N N L L H K L K I F H I S Q S E D K A N I L D K D
 E H F Y L V F E E C Y F E L A N I V P L Y N K I R N Y I T Q K P Y S D E K F K L N F E N S T L A N G W D K N K E P D N T A I
 L F I K D D K Y Y L G V M N K K N N K I F D D K A I K E N K G E G Y K K I V Y K L L P G A N K M L P K V F F S A K S I K F Y
 30 N P S E D I L R I R N H S T H T K N G S P Q K G Y E K F E F N I E D C R K F I D F Y K Q S I S K H P E W K D F G F R F S D T
 Q R Y N S I D E F Y R E V E N Q G Y K L T F E N I S E S Y I D S V V N Q G K L Y L F Q I Y N K D F S A Y S K G R P N L H T L
 Y W K A L F D E R N L Q D V V Y K L N G E A E L F Y R K Q S I P K K I T H P A K E A I A N K N K D N P K K E S V F E Y D L I
 K D K R F T E D K F F F H C P I T I N F K S S G A N K F N D E I N L L L K E K A N D V H I L S I A R G E R H L A Y Y T L V D
 G K G N I I K Q D T F N I I G N D R M K T N Y H D K L A A I E K D R D S A R K D W K K I N N I K E M K E G Y L S Q V V H E I

AKLVIEYNAIVVF**A**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNILVFKDNEFDKTGGVLRA
 YQLTAPFETFKKMGKQTGI IYYVPAGFTSKICPVTGTFVNQLYPKYESVSKSQEFFSKFDKIC
 YNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLL
 KDYSIEYGHGECIKAAICGESDKKFFAKLTSLVNTILQMRNSKTGTELDYLISPADVNGNF
 5 FDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

Francisella novicida CpfI D917A/D1255A (A917, E1006, and A1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDDEKRAKDYKKAKQIIDKYHQFFI
 10 EEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFN
 QNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEI IKSFKGWTTYFKGFHENR
 KNVYSSNDIPTSI IYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDY
 KTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQ
 INDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTMQSFYEQIAAFKTVEEKSIKE
 15 TLLSLFDDLKAQKLDLSKIYFKNDKSLTDLQQVFDDYSVIGTAVLEYITQQIAPKNLDNPS
 KKEQELIAKKTEKAKYLSLETIKLAL EEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
 DNLAQISIKYQNOGKKDLLQASAEDDVKAIKDLLDQTNLLHKLKIFHISQSEDKANILDKD
 EHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAI
 LFIKDDKYYLGVNMKNKNNKIFDDKA IENKGEYKKIVYKLLPGANKMLPKVFFSAKSIKFY
 20 NPSEDILRIRNHSTHTKNGSPQGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDT
 QRYNSIDEFYREVENQGYKLT FENISESYIDSVVNQGLYLFQIYNKDFSAYSKGRPNLHTL
 YWKALFDERNLQDVVYKLNGEAE LFYRKQSI PKKITHPAKEAIANKNDNPKKESVFEYDLI
 KDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLEKEANDVHILSI**A**ARGERHLAYYTLDV
 GKGNI I KQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEI
 25 AKLVIEYNAIVVF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNILVFKDNEFDKTGGVLRA
 YQLTAPFETFKKMGKQTGI IYYVPAGFTSKICPVTGTFVNQLYPKYESVSKSQEFFSKFDKIC
 YNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLL
 KDYSIEYGHGECIKAAICGESDKKFFAKLTSLVNTILQMRNSKTGTELDYLISPADVNGNF
 FDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

30

Francisella novicida CpfI E1006A/D1255A (D917, A1006, and A1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDDEKRAKDYKKAKQIIDKYHQFFI
 EEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFN

QNLIDAKKGQESDLILWLKQSKDNGIELFKANS DITDIDEALEI IKSFKGWTTYFKGFHENR
 KNVYSSNDIPTSI IYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDY
 KTSEVNQRVFSLDEVFEIANFN NYLNQSGITKFNTI IGGKFVNGENTKRKGINEYINLYSQQ
 INDKTLKKYKMSVLFKQILSDTESKS FVIDKLEDDSDVVTMQSFYEQIAAFKTVEEKSIKE
 5 TLLSLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPS
 KKEQELIAKKTEKAKYLSLETIKLAL EEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
 DNLAQISIKYQNOGKKDLLQASAEDDVKA IKDLLDQTNLLHKLKIFHISQSEDKANILDKD
 EHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAI
 LFIKDDKY YLGVMNKKNNKIFDDKA IENKGEYKKIVYKLLPGANKMLPKVFFSAKSIKFY
 10 NPSEDILRIRNHSTHTKNGSPQKG YEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDT
 QRYNSIDEFYREVENQGYKLT FENISESYIDS VVNQGLYLFQIYNKDFSAYS SKGRPNLHTL
 YWKALFDERNLQDVVYKLNGEAE LFYRKQSI PKKITHPAKEAIANKNDNPKKESVFEYDLI
 KDKRFTEDKFFFHC PITINFKSSGANKFNDEINLLLEKANDVHILSI DRGERHLAYYTLDV
 GKGNI IKQDTFNI IGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEI
 15 AKLVIEYNAIVVF ADLNFGFKRGRFKVEKQVYQKLEKMLIEKLN YLVFKDNEFDKTGGVLRA
 YQLTAPFETFKKMGKQTGI IYYVPAGFTSKICPVTG FVNQLYPKYESVSKSQEFFSKFDKIC
 YNLDKGYFEFSFDYKNFGDKAAKGWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLL
 KDYSIEYGHGECIKAAICGESDKKFFAKLT SVLNTILQMRNSKTGTELDYLISP VADVNGNF
 FDSRQAPKNMPQDA AANGAYHIGLGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

20

Francisella novicida CpfI D917A/E1006A/D1255A (A917, A1006, and A1255 are bolded and underlined)

MSIQEFVNKYSLSKTLRFELIPQGKTLENIKARGLIL DDEKRAKDYKKAKQIIDKYHQFFI
 EEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSADTIKKQISEYIKDSEKFKNLFN
 25 QNLIDAKKGQESDLILWLKQSKDNGIELFKANS DITDIDEALEI IKSFKGWTTYFKGFHENR
 KNVYSSNDIPTSI IYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDY
 KTSEVNQRVFSLDEVFEIANFN NYLNQSGITKFNTI IGGKFVNGENTKRKGINEYINLYSQQ
 INDKTLKKYKMSVLFKQILSDTESKS FVIDKLEDDSDVVTMQSFYEQIAAFKTVEEKSIKE
 TLLSLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPS
 30 KKEQELIAKKTEKAKYLSLETIKLAL EEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
 DNLAQISIKYQNOGKKDLLQASAEDDVKA IKDLLDQTNLLHKLKIFHISQSEDKANILDKD
 EHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAI
 LFIKDDKY YLGVMNKKNNKIFDDKA IENKGEYKKIVYKLLPGANKMLPKVFFSAKSIKFY
 NPSEDILRIRNHSTHTKNGSPQKG YEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDT

QRYNSIDFYREVENQGYKLT FENISESYIDSVVNQGKLYLFQIYNKDFSAYS SKGRPNLHTL
 YWKALFDERNLQDVVYKLNGEAELFYRKQSI PKKITHPAKEAIANKNDNPKKESVFEYDLI
 KDKRFTEDKFFFFHCPITINFKSSGANKFNDEINLLLEKEANDVHILSI ARGERHLAYYTLDV
 GKGNI IKQDTFNI IGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEI
 5 AKLVIEYNAIVVF ADLNFGFKRGRFKVEKQVYQKLEKMLIEKLNLYLVFKDNEFDKTGGVLRA
 YQLTAPFETFKKMGKQTGI IYYVPAGFTSKICPVTGTFVNQLYPKYESVSKSQEFFSKFDKIC
 YNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLL
 KDYSIEYGHGECIKAAICGESDKKFFAKLT SVLNTILQMRNSKTGTELDYLISPVADVNGNF
 FDSRQAPKNMPQDA AANGAYHIGLKLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

10 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 domain is a Cas9 domain from *Staphylococcus aureus* (SaCas9). In some embodiments, the SaCas9 domain is a nuclease active SaCas9, a
 15 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 herein.

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
 20 embodiments, the SaCas9 domain, the SaCas9d domain, or the SaCas9n domain can bind to a nucleic acid sequence having a NNGRRT or a NNGRRT 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
 25 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.

Exemplary SaCas9 sequence

30 KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHR
 IQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVE
 EDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTS DYVKEAKQLLKVQ
 KAYHQLDQSFIDTYIDLLETRRTYYEGPGE GSPFGWKDIKEWYEMLMGHCTYFPEELRSVKY

AYNADLYNALNDLNNLVITRDENEKLEYEYEFQIIENVFKQKKKPTLKQIAKEILVNEEDIK
 GYRVTSTGKPEFTNLKVYHDIKDITARKEIENAELLDQIAKILTIYQSSEDIQEELTNLNS
 ELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKE
 IPTTLVDDFILSPVVKRSFIQSIKVINAI IKKYGLPNDII IELAREKNSKDAQKMINEMQKR
 5 NRQTNERIEEII RTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHII
 PRSVSFDNSFNNKVLVKQEE**N**SKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISK
 KKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLRSYFRVNNLDVKVKSINGGFTSF
 LRRKWKFKKERNKGYKHHAE DALI IANADFIKWKKLDKAKKVMENQMFE EKQAESMPEIE
 TEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNG
 10 LYDKDNDKLKLINKSPEKLLMYHHPQTYQKLKLIMEQYGDEKNPLYKYEEETGNYLTKYS
 KKDNGPVIKKIKYYGNKLNALHDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNL
 DVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDDLNRIE
 VN MIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG

15 Residue N579 above, which is underlined and in bold, may be mutated (*e.g.*, to a
 A579) to yield a SaCas9 nickase.

Exemplary SaCas9n sequence

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHR
 20 IQRVKLLFDYNLLTDHSELSGINPYEARVKGLS QKLSEEEFSAALLHLAKRRGVHNVNEVE
 EDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFTSDYVKEAQQLLKVQ
 KAYHQLDQS FIDTYIDLLETRRTY YEGPGE GSPFGWKDIKEWYEMLMGHCTYFPEELRSVKY
 AYNADLYNALNDLNNLVITRDENEKLEYEYEFQIIENVFKQKKKPTLKQIAKEILVNEEDIK
 GYRVTSTGKPEFTNLKVYHDIKDITARKEIENAELLDQIAKILTIYQSSEDIQEELTNLNS
 25 ELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKE
 IPTTLVDDFILSPVVKRSFIQSIKVINAI IKKYGLPNDII IELAREKNSKDAQKMINEMQKR
 NRQTNERIEEII RTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHII
 PRSVSFDNSFNNKVLVKQEE**A**SKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISK
 KKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLRSYFRVNNLDVKVKSINGGFTSF
 30 LRRKWKFKKERNKGYKHHAE DALI IANADFIKWKKLDKAKKVMENQMFE EKQAESMPEIE
 TEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNG
 LYDKDNDKLKLINKSPEKLLMYHHPQTYQKLKLIMEQYGDEKNPLYKYEEETGNYLTKYS
 KKDNGPVIKKIKYYGNKLNALHDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNL
 DVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDDLNRIE
 35 VN MIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG

Residue A579 above, which can be mutated from N579 to yield a SaCas9 nickase, is underlined and in bold.

Exemplary SaKKH Cas9

5 K R N Y I L G L D I G I T S V G Y G I I D Y E T R D V I D A G V R L F K E A N V E N N E G R R S K R G A R R L K R R R R H R
I Q R V K K L L F D Y N L L T D H S E L S G I N P Y E A R V K G L S Q K L S E E E F S A A L L H L A K R R G V H N V N E V E
E D T G N E L S T K E Q I S R N S K A L E E K Y V A E L Q L E R L K K D G E V R G S I N R F K T S D Y V K E A K Q L L K V Q
K A Y H Q L D Q S F I D T Y I D L L E T R R T Y Y E G P G E G S P F G W K D I K E W Y E M L M G H C T Y F P E E L R S V K Y
10 A Y N A D L Y N A L N D L N N L V I T R D E N E K L E Y Y E K F Q I I E N V F K Q K K K P T L K Q I A K E I L V N E E D I K
G Y R V T S T G K P E F T N L K V Y H D I K D I T A R K E I I E N A E L L D Q I A K I L T I Y Q S S E D I Q E E L T N L N S
E L T Q E E I E Q I S N L K G Y T G T H N L S L K A I N L I L D E L W H T N D N Q I A I F N R L K L V P K K V D L S Q Q K E
I P T T L V D D F I L S P V V K R S F I Q S I K V I N A I I K K Y G L P N D I I I E L A R E K N S K D A Q K M I N E M Q K R
N R Q T N E R I E E I I R T T G K E N A K Y L I E K I K L H D M Q E G K C L Y S L E A I P L E D L L N N P F N Y E V D H I I
15 P R S V S F D N S F N N K V L V K Q E E **A** S K K G N R T P F Q Y L S S S D S K I S Y E T F K K H I L N L A K G K G R I S K T
K K E Y L L E E R D I N R F S V Q K D F I N R N L V D T R Y A T R G L M N L L R S Y F R V N N L D V K V K S I N G G F T S F
L R R K W K F K K E R N K G Y K H H A E D A L I I A N A D F I F K E W K K L D K A K K V M E N Q M F E E K Q A E S M P E I E
T E Q E Y K E I F I T P H Q I K H I K D F K D Y K Y S H R V D K K P N R K L I N D T L Y S T R K D D K G N T L I V N N L N G
L Y D K D N D K L K K L I N K S P E K L L M Y H H D P Q T Y Q K L K L I M E Q Y G D E K N P L Y K Y Y E E T G N Y L T K Y S
20 K K D N G P V I K K I K Y Y G N K L N A H L D I T D D Y P N S R N K V V K L S L K P Y R F D V Y L D N G V Y K F V T V K N L
D V I K K E N Y Y E V N S K C Y E E A K K L K K I S N Q A E F I A S F Y K N D L I K I N G E L Y R V I G V N N D L L N R I E
V N M I D I T Y R E Y L E N M N D K R P P H I K T I A S K T Q S I K K Y S T D I L G N L Y E V K S K K H P Q I I K K G.

Residue A579 above, which can be mutated from N579 to yield a SaCas9 nickase, is underlined and in bold. Residues K781, K967, and H1014 above, which can be mutated from
25 E781, N967, and R1014 to yield a SaKKH Cas9 are underlined and in italics.

In some embodiments, the napDNAbp is a circular permutant. In the following sequences, 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, and double underlined
30 sequence indicates mutations.

CP5 (with MSP “NGC” PID and “D10A” nickase):

E I G K A T A K Y F F Y S N I M N F F K T E I T L A N G E I R K R P L I E T N G E T G E I V W D K G R D F A T V R K V L S M
P Q V N I V K K T E V Q T G G F S K E S I L P K R N S D K L I A R K K D W D P K K Y G G F M Q P T V A Y S V L V V A K V E K

GKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKDLIIKLPKYSLELENGRKRML
 LASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISE
 FSKRVILADANLDKVL SAYNKH RD KPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYR
 STKEVL DATLIHQSI TGLYETRIDLSQLGGDGGSGGSGGSGGSGGSGGSGGMDKKYSIGLAI
 5 GTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
 RRKNRICYLQEIFS NEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
 YHLRKKLV DSTKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLF
 ENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLA
 EDAKLQLSKDTYDDDLNLLAQIGDQYADLF LAAKNLSDAILLSDILRVNTEITKAPLSASM
 10 IKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKM
 DGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPLKDNREKIEKILT
 FRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASAQSFIERMTNFDKNLPNEKV
 LPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
 KKIECFDSVEISGVEDRFNASLGTYHDLKI IKDKDFLDNEENEDILEDIVLTTLTFEDREM
 15 IEERLKTIAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
 MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHK
 PENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
 NGRDMYVDQELDINRLSDYVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNPSEEVVKKM
 KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
 20 KYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVG TALIKKYPK
 LESEFVYGDYKVYDVRKMI AKSEQEGADKRTADGSEFESPKKKRKV*

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
 25 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
 30 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.

The crystal structure of *Alicyclobacillus 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 least 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 least 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.

A Cas12b/C2c1 ((uniprot.org/uniprot/T0D7A2#2) sp|T0D7A2|C2C1_ALIAG CRISPR-associated endonuclease C2c1 OS = *Alicyclobacillus acido-terrestris* (strain ATCC

49025 / DSM 3922/ CIP 106132 / NCIMB 13137/GD3B) GN=c2c1 PE=1 SV=1) amino acid sequence is as follows:

MAVKS IKVKLRLLDDMP EIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQECD
 KTAEECKAELLERLRARQVENGHGPGAGSDDELLQLARQLYELLVPQAIGAKGDAQQIARKF
 5 LSPLADKDAVGGLGIAKAGNKPRWVRMREAGEPGWEEEEKAEATRK SADRTADVLRALADFG
 LKPLMRVYTDSEMSVVEWKPLRKGQAVRTWDRDMFQQA IERMMSWESWNQRVGQEYAKLVEQ
 KNRFEQKNFVGQEHLVHLVNQLQQDMKEASPGLESKEQTAHYVTGRALRGSDKVFEEKWGKLA
 PDAPFDLYDAEIKNVQRRNTRRFGSHDLFAKLAEPEYQALWREDASFLTRYAVYNSILRKLN
 HAKMFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGERRHAIRFHKLLKVENGVA REV
 10 DDVTVPISMSEQLDNLLPRDPNEPIALYFRDYGAEQHFTGEFGGAKIQCRRDQLAHMHRRRG
 ARDVYLNVSVRVQSQSEARGERRPPYAAVFRLVGDNHRA FVHFDKLSDYLAEHPDDGKLGSE
 GLLSGLRVMSVDLGLRTSASISVFRVARKDELKPN SKGRVPFFFFPIKGNDNLVAVHERS QLL
 KLPGETESKDLRAIREERQRTLRLQLRTQLAYLRLLLVRCGSEDVGRRERSWAKLIEQPVD AAN
 HMTPDWREAFENELQKLKSLHGICSDKEWMDAVYESVRRVWRHMGKQVRDWRKDVRS GERPK
 15 IRGYAKDVVGNSIEQIEYLERQYKFLKSWSF FGKVSQVIRAEKGSRFITLREHIDHAK E
 DRLKKLADRIIMEALGYVYALDERGKGKWKVAKYPPCQLILLEELSEYQFNNDRPPSENNQ L M
 QWSHRGVFQELINQAQVHDLLVGTMYAAFSSRF DARTGAPGIRCRRVPARCTQEHNPEFPFW
 WLNKFVVEHTLDACPLRADDLIPTGEGEIFVSPFSAEEGDFHQIHADLNAAQN LQQRLWSDF
 DISQIRLRCDWGEVDGELVLI PRLTGKRTADSYSNKV FYTNTGVTYERERERGGKRRKVFAQE
 20 KLSEEEAEELLVEADEAREKSVVLMRDP SGIINRGNWTRQKEFWSMV NQRIEGLVKQIRSR
 VPLQDSACENTGDI

AacCas12b (*Alicyclobacillus acidiphilus*) - WP_067623834

MAVKS MKVKLRLLDNMP EIRAGLWKLHTEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQECY
 25 KTAEECKAELLERLRARQVENGHGPGAGSDDELLQLARQLYELLVPQAIGAKGDAQQIARKF
 LSPLADKDAVGGLGIAKAGNKPRWVRMREAGEPGWEEEEKAKAEARKSTDR TADVLRALADFG
 LKPLMRVYTDSDMSSVQWKPLRKGQAVRTWDRDMFQQA IERMMSWESWNQRVGEAYAKLVEQ
 KSRFEQKNFVGQEHLVQLVNQLQQDMKEASHGLESKEQTAHYLTGRALRGSDKVFEEKWEKLD
 PDAPFDLYDTEIKNVQRRNTRRFGSHDLFAKLAE PKYQALWREDASFLTRYAVYNSIVRKLN
 30 HAKMFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGEGRHAIRFQKLLTVEDGVAKEV
 DDVTVPISM SAQLDDLLPRDPHELVALYFQDYGAEQHLAGEFGGAKIQYRRDQLNHLHARRG
 ARDVYLNLSVRVQSQSEARGERRPPYAAVFRLVGD NHRA FVHFDKLSDYLAEHPDDGKLGSE
 GLLSGLRVMSVDLGLRTSASISVFRVARKDELKPN SEGRVPFCFPIEGNENLVAVHERS QLL
 KLPGETESKDLRAIREERQRTLRLQLRTQLAYLRLLLVRCGSEDVGRRERSWAKLIEQPM DANQ

MTPDWREAFEDQLKLSLYGICGDREWTEAVYESVRRVVRHMGKQVRDWRKDVRSGERPKI
 RGYQKDVVGGNSIEQIEYLERQYKFLKSWFFGKVSQVIRAEKGSRAITLREHIDHAKED
 RLKKLADRIIMEALGYVYALDDERGGKGVAKYPPCQLILLEELSEYQFNDRPPSENNQLM
 QWSHRGVFQELLNQAQVHDLVGTMYAAFSSRFDARTGAPGIRCRVPARCAREQNPEFPW
 5 WLNKFVAEHKLDGCPLRADDLIPTGEGEFFVSPFSAEEGDFHQIHADLNAAQNLQRRLWSDF
 DISQIRLRCDWGEVDGEPVLIPTTGKRTADSYGNKVFTKTGVTYERERGGKRRKVFAQE
 ELSEEEAELLVEADEAREKSVVLMRDPGSIINRGDWTRQKEFWSMVNQRIEGLVKQIRSRV
 RLQESACENTGDI

10 **BhCas12b (*Bacillus hisashii*) NCBI Reference Sequence: WP_095142515**

MAPKKKRKVGIGHPAAATRSFILKIEPNEEVKGLWKTHEVLNHGIAAYMNLKLRQEI
 YEHHEQDPKNPKKVSKEIQAEWDFVLKMQKNSFTHEVDKDEVFNILRELYEELVPSSVE
 KKGEANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEKKKWEEDKKKDP
 LAKILGKLAEYGLIPLFIPTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWES
 15 WNLKVKEEYKVEKEYKTLEERIKEDIQALKALEQYEKERQEQLRDTLNTNEYRLSKRGLR
 GWREIIQKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPY
 LYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKL
 TVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKKGHAFTYKDESIFPLKGT
 LGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKP
 20 KELTEWIKDSKGKKLKGIESLEIGLRVMSIDLQGRQAAAASIFEVVDQKPDIEGKLFFPIK
 GTELYAVHRASFNIKLPGETLVKSREVLKAREDNLKLNMNQLNFLRNVLHFQQFEDITERE
 KRVTKWISRQENS DVPLVYQDELIQIRELMYKPYKDWVAFKQLHKLREVEIGKEVKHWRKS
 LSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFADQLNHLNALKED
 RLKKMANTIIMHALGYCYDVRKKKQAKNPACQIILFEDLSNYPYEERSRFENSKLMKWSR
 25 REIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFKLNQREGR
 LTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCK
 AYQVDGQTVYIPESKDQKQKIIEEFGEYFILKDGVEWVWVAGKLIKKGSSKQSSSELVDS
 DILKDSFDLASELKGEKLMLYRDPGSGNVFSPDKWMAAGVFFGKLERILISKLTNQYSISTIE
 DDSSKQSMKRPAATKKAGQAKKKK

30 In some embodiments, the Cas12b is BvCas12b (V4), which is a variant of BhCas12b and comprises the following changes relative to BhCas12b: S893R, K846R, and E837G.

BhCas12b (V4) is expressed as follows: 5' mRNA Cap---5'UTR---bhCas12b---STOP sequence --- 3'UTR --- 120polyA tail.

5'UTR:

GGGAAATAAGAGAGAAAAGAAGAGTAAGAAGAAATATAAGAGCCACC

3' UTR (TriLink standard UTR)

GCTGGAGCCTCGGTGGCCATGCTTCTTGCCCCTTGGGCCTCCCCCAGCCCCCTCCTCCCCTT
CCTGCACCCGTACCCCGTGGTCTTTGAATAAAGTCTGA

5 Nucleic acid sequence of bhCas12b (V4)

ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCCACCAGATC
CTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAAACCCACGAGG
TGCTGAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGGCAAGAGGCCATC
TACGAGCACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAGGCCGAGATCCAGGC
10 CGAGCTGTGGGATTTTCGTGCTGAAGATGCAGAAGTGCAACAGCTTCACACACGAGGTGGACA
AGGACGAGGTGTTCAACATCCTGAGAGAGCTGTACGAGGAACTGGTGCCAGCAGCGTGGA
AAGAAGGGCGAAGCCAACCAGCTGAGCAACAAGTTTCTGTACCCTCTGGTGGACCCCAACAG
CCAGTCTGGAAAGGGAACAGCCAGCAGCGGCAGAAAGCCCAGATGGTACAACCTGAAGATTG
CCGGCGATCCCTCCTGGGAAGAAGAGAAGAAGAAGTGGGAAGAAGATAAGAAAAAGGACCCG
15 CTGGCCAAGATCCTGGGCAAGCTGGCTGAGTACGGACTGATCCCTCTGTTTCATCCCCTACAC
CGACAGCAACGAGCCCATCGTGAAAGAAATCAAGTGGATGGAAAAGTCCCGGAACCAGAGCG
TGCGGCGGCTGGATAAGGACATGTTTCATTTCAGGCCCTGGAACGGTTCCTGAGCTGGGAGAGC
TGGAACCTGAAAGTGAAAGAGGAATACGAGAAGGTGAGAAAGAGTACAAGACCCTGGAAGA
GAGGATCAAAGAGGACATCCAGGCTCTGAAGGCTCTGGAACAGTATGAGAAAAGAGCGGCAAG
20 AACAGCTGCTGCGGGACACCCTGAACACCAACGAGTACCGGCTGAGCAAGAGAGGCCTTAGA
GGCTGGCGGGAAATCATCCAGAAATGGCTGAAAATGGACGAGAACGAGCCCTCCGAGAAGTA
CCTGGAAGTGTTCAAGGACTACCAGCGGAAGCACCCCTAGAGAGGCCGGCGATTACAGCGTGT
ACGAGTTCCTGTCCAAGAAAGAGAACCCTTCATCTGGCGGAATCACCCCTGAGTACCCCTAC
CTGTACGCCACCTTCTGCGAGATCGACAAGAAAAAGAAGGACGCCAAGCAGCAGGCCACCTT
25 CACACTGGCCGATCCTATCAATCACCCCTCTGTGGGTCCGATTGAGGAAAGAAGCGGCAGCA
ACCTGAACAAGTACAGAATCCTGACCGAGCAGCTGCACACCGAGAAGCTGAAGAAAAAGCTG
ACAGTGCAGCTGGACCGGCTGATCTACCCTACAGAATCTGGCGGCTGGGAAGAGAAGGGCAA
AGTGGACATTGTGCTGCTGCCCAGCCGGCAGTTCTACAACCAGATCTTCCTGGACATCGAGG
AAAAGGGCAAGCACGCCTTCACCTACAAGGATGAGAGCATCAAGTTCCTCTGAAGGGCACA
30 CTCGGCGGAGCCAGAGTGCAGTTCGACAGAGATCACCTGAGAAGATACCCTCACAAGGTGGA
AAGCGGCAACGTGGGCAGAATCTACTTCAACATGACCGTGAACATCGAGCCTACAGAGTCCC
CAGTGTCCAAGTCTCTGAAGATCCACCGGGACGACTTCCCCAAGGTGGTCAACTTCAAGCCC
AAAGAACTGACCGAGTGGATCAAGGACAGCAAGGGCAAGAACTGAAGTCCGGCATCGAGTC
CCTGGAAATCGGCCTGAGAGTGATGAGCATCGACCTGGGACAGAGACAGGCCGCTGCCGCCT

CTATTTTCGAGGTGGTGGATCAGAAGCCCGACATCGAAGGCAAGCTGTTTTTCCCAATCAAG
 GGCACCGAGCTGTATGCCGTGCACAGAGCCAGCTTCAACATCAAGCTGCCCCGGCGAGACACT
 GGTC AAGAGCAGAGAAGTGCTGCGGAAGGCCAGAGAGGACAATCTGAAACTGATGAACCAGA
 AGCTCAACTTCCTGCGGAACGTGCTGCACTTCCAGCAGTTCGAGGACATCACCGAGAGAGAG
 5 AAGCGGGTCACCAAGTGGATCAGCAGACAAGAGAACAGCGACGTGCCCCCTGGTGTACCAGGA
 TGAGCTGATCCAGATCCGCGAGCTGATGTACAAGCCTTACAAGGACTGGGTGCGCTTCCTGA
 AGCAGCTCCACAAGAGACTGGAAGTCGAGATCGGCAAAGAAGTGAAGCACTGGCGGAAGTCC
 CTGAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAGAACATCGACGAGATCGATCG
 GACCCGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACCGAACCTGGCGAAGTGCGTAGAC
 10 TGGAACCCGGCCAGAGATTCGCCATCGACCAGCTGAATCACCTGAACGCCCTGAAAGAAGAT
 CGGCTGAAGAAGATGGCCAACACCATCATCATGCACGCCCTGGGCTACTGCTACGACGTGCG
 GAAGAAGAAATGGCAGGCTAAGAACCCCGCCTGCCAGATCATCCTGTTGAGGATCTGAGCA
 ACTACAACCCCTACGAGGAAAGGTCCCGCTTCGAGAACAGCAAGCTCATGAAGTGGTCCAGA
 CGCGAGATCCCCAGACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTGGGAGAAGT
 15 GGGCGCTCAGTTCAGCAGCAGATTCCACGCCAAGACAGGCAGCCCTGGCATCAGATGTAGCG
 TCGTGACCAAAGAGAAGCTGCAGGACAATCGGTTCTTCAAGAATCTGCAGAGAGAGGGCAGA
 CTGACCCTGGACAAAATCGCCGTGCTGAAAGAGGGCGATCTGTACCCAGACAAAGGCGGCGA
 GAAGTTCATCAGCCTGAGCAAGGATCGGAAGTGCCTGACCACACACGCCGACATCAACGCCG
 CTCAGAACCTGCAGAAGCGGTTCTGGACAAGAACCCACGGCTTCTACAAGGTGTACTGCAAG
 20 GCCTACCAGGTGGACGGCCAGACCGTGTACATCCCTGAGAGCAAGGACCAGAAGCAGAAGAT
 CATCGAAGAGTTTCGGCGAGGGCTACTTCATTCTGAAGGACGGGGTGTACGAATGGGTCAACG
 CCGGCAAGCTGAAAATCAAGAAGGGCAGCTCCAAGCAGAGCAGCAGCGAGCTGGTGGATAGC
 GACATCCTGAAAGACAGCTTCGACCTGGCCTCCGAGCTGAAAGGCGAAAAGCTGATGCTGTA
 CAGGGACCCAGCGGCAATGTGTTCCCCAGCGACAAATGGATGGCCGCTGGCGTGTTCTTCG
 25 GAAAGCTGGAACGCATCCTGATCAGCAAGCTGACCAACCAGTACTCCATCAGCACCATCGAG
 GACGACAGCAGCAAGCAGTCTATGAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAA
 AAAGAAAAAG

In some embodiments, the Cas12b is BvCas12B. In some embodiments, the Cas12b
 comprises amino acid substitutions S893R, K846R, and E837G as numbered in the
 30 BvCas12b exemplary sequence provided below.

BvCas12b (*Bacillus* sp. V3-13) NCBI Reference Sequence: WP_101661451.1

MAIRSIKLMKTNSGTDSIYLRKALWRTHQLINEGIAYMNLTLTYRQEAIQDKTKEAYQAE
 LINIIRNQQRNNGSSEEHGSDQEILALLRQLYELIIPSSIGESGDANQLGNKFLYPLVDPNS
 QSGKGTSNAGRKPRWKRLKEEGNPDWELEKKKDEERKAKDPTVKIFDNLNKYGLLPLFPLFT

NIQKDIEWLPLGKRQSVRKWDKDMFIQAIERLLSWESWNRRVADEYKQLKEKTESYYKEHLT
 GGEWIEKIRKFEKERNMELEKNAFAPNDGYFITSRQIRGWDRVYEKWSKLPESASPEELWK
 VVAEQQNMSEGFGDPKVSFLANRENDRDIWRGHSEIRYHIAAYNGLQKKLSRTKEQATFTL
 PDAIEHPLWIRYESPGGTNNLNFKLEEKQKKNYVTLISKIWPSEEKWKIEKENIEIPLAPSI
 5 QFNRQIKLKQHVKGKQEISFSDYSSRISLDGVLGGSRIQFNRKYIKNHKELLGEGDIGPVFF
 NLVVDVAPLQETRNGRLQSPIGKALKVISSDFSKVIDYKPKELMDWMNTGSASNSFGVASLL
 EGMRVMSIDMGQRTSASVSI FEVVKELPKDQEQKLFYSINDTELF AIHKRSFLLNLPGEVVT
 KNNKQQRQERRKKRQFVRSQIRMLANVLRLETKKT PDERKKA I HKLMEIVQS YDSWTASQKE
 VWEKELNLLTNMAAFNDEIWKESLVELHHRIEYPVGQIVSKWRKGLSEGRKNLAGISMWNID
 10 ELEDTRRL LISWSKRSRTPGEANRIETDEPFGSSLLQHIQNVKDDRLKQMANLI IMTALGFK
 YDKEEK DRYKRWKETYPACQIILFENLNRYLFNLDRSRRENSRLMKWAHRSI PRTVSMQ GEM
 FGLQVG D VRSEYSSRFHAKTGAPGIRCHALTEEDLKAGSNTLKR LIEDGFINESELAYLKKG
 DIIPSQGGELFVTLSKRYKKDS DNNE LTVIHADINAAQNLQKRFWQQNSEVYRVPCQLARMG
 EDKLYI PKSQTTETIKKYFGKGSFVKNNTEQEVYKWEKSEMKIKTDTT FDLQDLG FEDISK
 15 TIELAQEQKKYLTMFRDPSGYFFNNETWRPQKEYWSIVNNI IKSC LKKKILSNKVEL

In some embodiments, the Cas12b is **BTCas12b**. **BTCas12b** (*Bacillus*

thermoamylovorans) NCBI Reference Sequence: WP_041902512

MATRSFILKIEPNEEVKKGLWKTHEVLNHGIAAYMNI LKLI RQEAIYEHHEQDPKNPKKV
 SKAEIQ AELWDFVLKMQKCNSFTHEVDKDVFNILRELYEELVPSSVEKKGEANQLSNKF
 20 LYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEKKKWEEDKKKDPLAKILGKLAE
 YGLIPLFIPFTDSNEPIVKEIKWMEKSRNQSVRR LDKDMFIQALERFLSWESWNLKVKEE
 YEKVEKEHKTLEERIKEDIQAFKSLEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREII
 QKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYAT
 FCEIDKKKKDAKQQATFTLADP INHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLT V
 25 QLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDES IKFPLKGT
 LGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKFVNF
 KPKELTEWIKDSKGKKLKSGIESLEIGLRVMSIDL GQRQAAAASIFEVVDQKPDIEGKLF
 FPIKGT ELYAVHRASFNIKLPGETLVKSREVL RKAREDNLKL MNQKLNFLRNVLHFQQFE
 DITEREKRVTKWISRQENS DVPLVYQDELIQIRELMYKPYKDWVAF LKQLH KRLEVEIGK
 30 EVKHWKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQ
 LNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWKQAKNPACQIILFEDLSNYPYEERS
 RFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKL
 QDNRF FKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKLVTTHADINAAQNLQ
 KRFWTRTHGFYKVYCKAYQVDGQTVYI PESKDQKQKII EEFGE GYFILKDG VYEWGNAGK

LKIKKGSSKQSSSELVDS DILKDSFDLASELKGEKLM LYRDP SGNVFPSDKWMAAGVFFG
KLERILISKLTNQYSISTIEDDSSKQSM

In some embodiments, a napDNAbp refers to Cas12c. In some embodiments, the Cas12c protein is a Cas12c1 or a variant of Cas12c1. In some embodiments, the Cas12 protein is a Cas12c2 or a variant of Cas12c2. In some embodiments, the Cas12 protein is a Cas12c protein from *Oleiphilus* sp. HI0009 (*i.e.*, OspCas12c) 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 least 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.

Cas12c1

MQTKKTHLHLISAKASRKYRRTIACLSDTAKKDLERRKQSGAADPAQELSCLKTIKFKLEVP
EGSKLP SFDRISQIYNALETIEKGSLSYLLFALILSGFRI FPNSSAAKTFASSSCYKNDQFA
SQIKEIFGEMVKNFIPSELESILKKGRRKNNKDWTEENIKRVLNSEFGRKNSEGSSALFDSF
LSKFSQELFRKFDSWNEVNKKYLEAAELLD SMLASYGPFDSVCKMIGDSDSRNSLPDKSTIA
FTNNAEITVDIESSVMPYMAIAALLREYRQSKSKAAPVAYVQSHLT TTNGNGLSWFFKFGLD
LIRKAPVSSKQSTSDGSKSLQELFSVPDDKLDGLKFIKEACEALPEASLLCGEKGELLGYQD
FRTSFAGHIDSWVANYVNRLFELIELVNQLPESIKLPSILTQKNHNLVASLGLQEAEVSHSL
ELFEGLVKNV RQTLKKLAGIDISSPNEQDIKEFYAFSDVLNRLGSIRNQIENAVQTAKKDK
IDLESAIEWKEWKKLKKLPKLNLGGGVPKQQELLDKALESVKQIRHYQRIDFERVIQWAVN
EHCLETVPKFLVDAEKKKINKESSTDFAAKENAVRFLLEGIGAAARGKTDSVSKAAYNWFV
NNFLAKKDLNRYFINCQGCYKPPYSKRSLAFALRSDNKDTIEVVWEKFETFYKEISKEIE
KFNIFSQEFQTFHLLENLRMKLLLRRIQKPIPAEIAFFSLPQEYYDSLPPNVAFLALNQEIT
PSEYITQFNLYSSFLNGNLILLRRSRSYLRAKFSWVGNSKLIYAAKEARLWKIPNAYWKSDE
WKMILDSNVLVFDKAGNVLPAPTLKKVCEREGDLRLFYPLL RQLPHDWCYRNP FVKSVGREK

NVIEVNKEGEPKVASALPGSLFRLIGPAPFKSLDDCFNPLDKDLRECMLIVDQEISQKVE
 AQKVEASLESCTYSIAVPIRYHLEEPKVSNQFENVLAIDQGEAGLAYAVFSLKSIGEAETKP
 IAVGTIRIPSIIRLIHSVSTYRKKKQRLQNFQNYDSTAFIMRENTGVDCAKIVGLMKEFN
 AFPVLEYDVKNLESGSRQLSAVYKAVNSHFLYFKEPGRDALRKQLWYGGDSWTIDGIEIVTR
 5 ERKEDGKEGVEKIVPLKVFPGRSVSARFTSKTSCCCGRNVFDWLFTEKKAKTNKKFNVNSKG
 ELTTADGVIQLFEADRSKGPKFYARRKERTPLTKPIAKGSYSLEEIERVRNLRAPKSKQ
 SRDTSQSQYFCVYKDCALHFSGMQADENAAINIGRRFLTALRKNRRSDFPSNVKISDRLLDN
 Cas12c2

MTKHSIPLHAFRNSGADARKWKGRIALLAKRGETMRTLQFPLEMSEPEAAAINTTPFAVAY
 10 NAIEGTGKGTLDYWAKLHLAGFRFFPSGGAATIFRQQAVFEDASWNA AFCQQSGKDWPWL
 PSKLYERFTKAPREVAKKDGSKKSIEFTQENVANESHVSLVGASITDKTPEDQKEFFLK
 ALAEKFDSWKSANEDRIVAMKVIDEFLKSEGLHLPSLENI AVKCSVETKPDNATVAWHDAPM
 SGVQNLAIGVFATCASRIDNIYDLNGGKLSKLIQESATTPNVTALSWLFGKGLEYFRTT
 TIMQDFNIPASAKESIKPLVESQAIPMTVLGKKNYAPFRPNFGGKIDSWIANYASRLMLL
 15 NDILEQIEPGFELPQALLDNETLMSGIDMTGDELKELIEAVYAWVDAAKQGLATLLGRGGN
 DDAVQTFEQFSAMMDTLNGTLNTISARYVRVEMAGKDEARLEKLECKFDIPKWKSVPKL
 VGISGGLPKVEEEIKVMNAAFKDVRARMFVRFEETIAAYVASKGAGMDVYDALEKRELEQIKK
 LKS AVPERAHIQAYRAVLHRIGRAVQNCSEKTKQLFSSKVIEMGVFKNPSHLNFI FNQKGA
 IYRSPFDRSRHAPYQLHADKLLKNDWLELLAEISATLMASESTEQMEDALRLERTRLQLQLS
 20 GLPDWEYPASLAKPDIEVEIQ TALKMQLAKDTVTS DVLQRAFNLVSSVLSGLTFKLLRRSFS
 LKMRFSVADTTQLIYVPKVC DWAI PKQYLQAEGEIGIAARVVTESSPAKMVTEVEMKEPKAL
 GHFMQQAPHDWYFDASLGGTQVAGRIVEKGKEVGKERKLVGYRMRGNSAYKTVL DKS LVGNT
 ELSQCSMIEIPYTQTVDA DFRAQVQAGLPKVSINLPVKETITASNKDEQMLFDRFVAIDL
 ERGLGYAVFDAKTLELQESGHRPIKAITNLLNRTHHYEQRPNQ RQKFQAKFNVNLSELRENT
 25 VGDVCHQINRICAYYNAFPVLEYMVPDRLDKQLKSVYESVTNRYIWSSTDAHKSARVQFWLG
 GETWEHPYLKSAKDKKPLVLS PGRGASGKGTSQTCSCCGRNPFDLIKDMKPRAKIAVVDGKA
 KLENSELKLFERNLESKDDMLARRHRNERAGMEQPLTPGNYTVDEIKALLRANLRAPKNRR
 TKDTTVSEYHCVFSDCGKTMHADENAAVNIGGKFIADIEK

OspCas12c

30 MTKLRHRQKKLTHDWAGSKKREVLGSNGKLQNPLLM PVKKGVTEFRKA FSA YARATKGEMT
 DGRKNMFTHSFEPFKTKPSLHQCELADKAYQSLHSYLPGLAHFLLSAHALGFRI FSKSGEA
 TAFQASSKIEAYESKLASELACVDLSIQNL TISTLFNALTT SVRGKGEETSADPLIARFYTL
 LTGKPLSRDTQGPERDLAEVISRKIASSFGTWKEMTANPLQSLQFFEEELHALDANVSLSPA
 FDVLIKMNDLQGD LKNRTIVFDPDAPVFEYNAEDPADII IKLTARYAKEAVIKNQNVGNYVK

NAITTTNANGLGWLLNKGLSLLPVSTDDLELLEFIGVERSHPSCHALIELIAQLEAPELFEKN
 VFSDTRSEVQGMIDSAVSNHIARLSSSRNSLSMDSEELERLIKSFQIHTPHCSLFIGAQSL
 QQLESLEALQSGVNSADILLGSTQYMLTNSLVEESIATYQRTLNRINYLSGVAGQINGAIK
 RKAIDGEKIHLPAAWSELISLPFIGQPVIDVESDLAHLKNQYQTLSENEFDTLISALQKNFDL
 5 NFNKALLNRTQHFEAMCRSTKKNALSKPEIVSYRDLLARLTSCLYRGSVLRRAGIEVLKKH
 KIFESNSELREHVHERKHFFVVSPLDRKAKKLLRLTDSRPDLLHVIDEILQHDNLENKDRES
 LWLVRSGYLLAGLPDQLSSSFINLPIITQKGDRLIDLIQYDQINRDAFVMLVTSAFKSNLS
 GLQYRANKQS FVVTRTLSPYLGSKL VYVPKDKDWLVPSQMFEGRFADILQSDYMWK DAGRL
 CVIDTAKHLSNIKKSVFSSEEVLAFRLRELPHRTFIQTEVRGLGVNVDGIAFNNGDIPSLKTF
 10 SNCVQVKVSRTNTSLVQTLNRWFEGGKVSPPSIQFERAYYKKDDQIHEDAARKIRFQMPAT
 ELVHASDDAGWTPSYLLGIDPGEYGMGLSLVSINNGEVLDSGFIHINSLINFASKKSNHQT
 VVPRQQYKSPYANYLEQSKDSAAGDIAHILDRLIYKLNALPVFEALSGNSQSAADQVWTKVL
 SFYTWGDNDNAQNSIRKQHWFGASHWDIKGMLRQPPTEKKPKPYIAFPGSQVSSYGNQRCSC
 CGRNPIEQLEMAKDTSIKELKIRNSEIQLFDTIKLFNPDPSTVIERRRHNLGPSRIPVAD
 15 RTFKNISPSSELEFKELITIVSRHSPEFIAKKRGIGSEYFCAYSDCNSSLNSEANAAANV
 AQKFQKQLFFEL

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
 20 by reference. 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
 25 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,
 30 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
 least 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.

Cas12g1

5 MAQASSTPAVSPRPRPRYREERTLVRKLLPRPGQSKQEFRENVKKLRKAFLQFNADVSGVCQ
WAIQFRPRYGKPAEPTETFWKFFLEPETS LPPNDSRSPEFRRLQAFEEAAGINGAAALDDPA
FTNELRDSILAVASRPKTKEAQR LFSRLKDYQPAHRMILAKVAAEWIESRYRRAHQNWERNY
EEWKKEKQEWEQNHPELTPEIREAFNQIFQQLEVKEKRVRICPAARLLQNKDNCQYAGKNKH
SVLCNQFNEFKKNHLQGKA IKFFYKDAEKYLRCLQSLKPNVQGPFRDWNKYLRYMNLKEE
10 TLRGKNGGRLPHCKNLGQECFNPHTALCKQYQQQLSSRPDLVQHDELYRKWRREYWREPRK
PVFRYPSVKRHSIAKIFGENYFQADFKNSVVGRLD SMPAGQYLEFAFAPWPRNYPQPGET
EISSVHLHFVGT RPRIGFRFRVPHKRSRFDCTQEELDELSRTFPRKAQDQKFLEAARKRL
ETFPGNAEQELRLLAVDLGTDSARAAFFIGKTFQQAFLKIVKIEKLYEQWPNQKQAGDRRD
ASSKQPRPGLSRDHVGRHLQKMRAQASEIAQKRQELTGT PAPETTTDQAAKKATLQPFDLRG
15 LTVHTARMIRDWARLNARQIIQLAEENQVDLIVLES LRGFPPGYENLDQEKRRVAFFAHG
RIRRKVTEKAVERGMRVVTPYLASSKVCAECRKKQKDNKQWEKNKKRGLFKCEGCGSQAQV
DENAARVLGRVFWGEIELPTAIP

Cas12h1

20 MKVHEIPRSQLLKIKQYEGSFVEWYRDLQEDRKKFASLLFRWAAFGYAAREDDGATYISPSQ
ALLERRLLLGD AEDVAIKFLDVLFKGGAPSSSCYSLFYEDFALRDKAKYSGAKREFIEGLAT
MPLDKI IERIRQDEQLSKI PAEEWLILGAEYSPEEIWEQVAPRIVNVDRSLGKQLRERLGIK
CRRPHDAGYCKILMEVVARQLRSHNETYHEYLNQTHEMKT KVANNLTNEFDLVCEFAEVLEE
KNYGLGWYVLWQGVKQALKEQKKPTKIQIAVDQLRQPKFAGLLTAKWRALKGAYDTWKLKKR
25 LEKRKAFFPYMPNWDNDYQIPVGLTGLGVFTLEV KRTEVVVDLKEHGKLFCSHSHYFGDLTAE
KHPSRYHLKFRHKLKLKRDSRVEPTIGPWIEAALREITIQKKPNGVFYLG LPYALSHGIDN
FQIAKRFFSAAKPDKEVINGLPSEMVG AADNLNSNIVAPVKARIGKGLEGPLHALDYGGE
LIDGPKILTPDGPRCGELISLKR DIVEIKSAIKEFKACQREGLTMSEETTTWLSEVESPSDS
PRCMIQSRIADTSRRLNSFKYQMNKEGYQDLAEALRLLDAMD SYNSLLESYQRMHLS PGEQS
30 PKEAKFDTKRASFRDLLRRRVAHTIVEYFDDCDIVFFEDLDG PSDSDSRNNALVKLLSPRTL
LLYIRQALEKRGIGMVEVAKDGTSQNNPISGHVGWRNKQNKSEIYFYEDKELLVMDADEVGA
MNILCRGLNHSVCPYSFVTKAPEKKNDEKKEGDYGKRVKRFLKDRYGSSNVRFLVASMGFVT
VTTKRPKDALVGKRLYYHG GELVTHDLHNRMKDEIKYLV EKEVLARRVSLSDSTIKSYKSFA
HV

Cas12i1

MSNKEKNASETRKAYTTKMI PRSHDRMKLLGNFMDYLM DGTPIFFELWNQFGGGIDRDIISG
 TANKDKISDDL LLA VNWFKVM PINSKPQGVSPSNLANLFQQYSGSEPDIQAQ EYFASNFDTE
 5 KHQWKDMRVEYERLLAELQLSRSDMHDLKLMYKEKCIGLSLSTAHYITSVMFGTGAKNNRQ
 TKHQFYSKVIQLLEESTQINSVEQLASII LKAGDCDSYRKLRI RCSRKGATPSILKIVQDYE
 LGTNHDDEVNVP SLIANLKEKLGRFEYECEWKMEKIKAF LASKVGPYYLGSYSAMLENALS
 PIKGMTTKNCKFVLKQIDAKNDIKYENEPFGKIVEGFFDSPYFESDTNVKWLHPHHIGESN
 IKTLWEDLNAIHSKYEEDIASLSEDKKEKRIKVYQGDVCQTINTYCEEVGKEAKTPLVQLLR
 10 YLYSRKDDIAVDKIIDGITFLSKKHKVEKQKINPVIQKYP SFNFGNNSKLLGKIISPDKDLK
 HNLKCNRNQVDNYIWI EIKVLNTKTMRWEKH HYALSSTRFLEEVYYPATSENPPDALAARFR
 TKTNGYEGKPALSAEQIEQIR SAPVGLRKVKKRQMRLEAARQQNLLPRYTWGKDFNINICKR
 GNNFEVTLATKVKKKKEKNYKVVLGYDANIVRKNTYAAIEAHANGDGVIDYNDLPVKPIESG
 FVTVESQVRDKSYDQLSYNGVKLLYCKPHVESRRSFLEKYRNGTMKDNRGNNIQIDFMKDFE
 15 AIADDETSLYYFNMKYCKLLQSSIRNHSSQAKEYREEIFELLRDGKLSVLKLSLSLNSL FVM
 FKVAKSLIGTYFGHLLKKPKNSKSDVKAPPITDEDKQKADPEMFALRLALEEKRLNKVKSKK
 EVIANKIVAKALELRDKYGPVLIKGENISDTTKKGKKSSTNSFLMDWLARGVANKVKEMVMM
 HQGLEFVEVNP NF TSHQDPFVHKNPENTFRARYSRCTPSELTEKNRKEILSFLSDKPSKRPT
 NAYYNEGAMAFLATYGLKKN DVLGVSLEKFKQIMANILHQ RSEDQLLFPSRGGMFYLATYKL
 20 DADATSVNWNGKQFWVCNADLVAAYNVGLVDIQKDFKKK

Cas12i2

MSSAIKSYKSVLRPNERNK NQLLKSTIQCLEDGSAFFFKMLQGLFGGITPEIVRFSTEQEKQQ
 QDIALWCAVNWFRPVSQDSLTH TIASDNLVEKFEEYYGGTASDAIKQYFSASIGESYYWNDC
 25 RQQYYDLCRELGVEVSDLTHDLEILCREKCLAVATESNQNSIISVLFGTGEKEDRSVKLRI
 TKKILEAISNLKEIPKNVAPIQEII LNVAKATKETFRQVYAGNLGAPSTLEKFI AKDGQKEF
 DLKKLQTDLKKVIRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRN
 YNFAKQRLEQFKEIQSLNNLLVVKLNDFFDSEFFSGEETY TICVHHLGGKDLSKLYKAWED
 DPADPENAIIVLCDDLKNNFKKEPIRNILRYIFTIRQECSAQDILAAAKYNQQLD RYKSQKA
 30 NPSVLGNQGFTWTNAVILPEKAQRNDRPNSLDLRIWL YLKL RHPDGRWKKHHIPFYDTRFFQ
 EIYAAGNSPVDTCCQFRTPRFGYHLPKLT DQTAIRVNKKHVKA AKTEARIRLAIQQGTLPVSN
 LKITEISATINSKGQVRIPVKFDVGRQKGT LQIGDRFCGYDQNTASHAYSLWEVVKEGQYH
 KELGC FVRFISSGDIVSITENRGNQFDQLSYEGLAYPQYADWRKKASKFVSLWQITKKNKKK
 EIVTVEAKEKFDAICKYQPRLYKFNKEYAYLLRDIVRGKSLVELQQIRQEIFRFIEQDCGVT

RLGSLSLSTLETVKAVKGI IYSYFSTALNASKNNPISDEQRKEFDPELFALLEKLELIRTRK
 KKQKVERIANSLIQTCLENNIKFIRGEGDLSTTNNATKKKANSRSMDWLARGVFNKIRQLAP
 MHNITLFGCGSLYTSHQDPLVHRNPDKAMKCRWAAIPVKDIGDWLRLKLSQNLRAKNIGTGE
 YYHQGVKEFLSHYELQDLEELLKWRSDRKSNI PCWVLQNRLAEKLGNEAVVYIPVRGGRI
 5 YFATHKVATGAVSIVFDQKQVWVCNADHVAAANIALTVKGIGEQQSSDEENPDGSRIKLQLTS

Representative nucleic acid and protein sequences of the base editors follow:

BhCas12b GSGGS-ABE8-Xten20 at P153

GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCCAC
 CAGATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAAACCC
 10 ACGAGGTGCTGAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGGCAAGAG
 GCCATCTACGAGCACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAGGCCGAGAT
 CCAGGCCGAGCTGTGGGATTTCTGTGCTGAAGATGCAGAAGTGCAACAGCTTCACACACGAGG
 TGGACAAGGACGAGGTGTTCAACATCCTGAGAGAGCTGTACGAGGAAGTGGTGCCACGACG
 GTGGAAGAAGGGCGAAGCCAACCAGCTGAGCAACAAGTTTCTGTACCCTCTGGTGGACCC
 15 CAACAGCCAGTCTGGAAAGGGAACAGCCAGCAGCGGCAGAAAGCCCAGATGGTACAACCTGA
 AGATTGCCGGCGATCCCgaggctctggaggaagcTCCGAAGTCGAGTTTCCCATGAGTAC
TGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGG
GGCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCC
ACGACCCCACTGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAAT
 20 TATCGACTTTATGATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGC
TATGATTCACTCCCGCATTGGACGAGTTGTATTGCGGTGTTGCAACGCCAAGACGGGTGCCG
CAGGTTCACTGATGGACGTGCTGCATCATCCAGGCATGAACCACCGGGTAGAAATCACAGAA
GGCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTGTTTTTTTCGCATGCCAGGCGGGT
CTTTAACGCCCAGAAAAAAGCACAAATCCTCTACTGACGGCTCTTCTGGATCTGAAACACCTG
 25 GCACAAGCGAGAGCGCCACCCCTGAGAGCTCTGGCTCCTGGGAAGAAGAGAAGAAGAAGTGG
 GAAGAAGATAAGAAAAAGGACCCGCTGGCCAAGATCCTGGGCAAGCTGGCTGAGTACGGACT
 GATCCCTCTGTTCATCCCCTACACCGACAGCAACGAGCCCATCGTGAAAGAAATCAAGTGGA
 TGGAAAAGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAAGGACATGTTTATTACAGGCCCTG
 GAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGTGAAAGAGGAATACGAGAAGGTCGA
 30 GAAAGAGTACAAGACCCTGGAAGAGAGGATCAAAGAGGACATCCAGGCTCTGAAGGCTCTGG
 AACAGTATGAGAAAGAGCGGCAAGAACAGCTGCTGCGGGACACCCTGAACACCAACGAGTAC
 CGGCTGAGCAAGAGAGGCCTTAGAGGCTGGCGGGAAATCATCCAGAAATGGCTGAAAATGGA
 CGAGAACGAGCCCTCCGAGAAGTACCTGGAAGTGTTCAAGGACTACCAGCGGAAGCACCCCTA
 GAGAGGCCGGCGATTACAGCGTGTACGAGTTCCTGTCCAAGAAAGAGAACCCTTCATCTGG

CGGAATCACCCCTGAGTACCCCTACCTGTACGCCACCTTCTGCGAGATCGACAAGAAAAAGAA
 GGACGCCAAGCAGCAGGCCACCTTCACACTGGCCGATCCTATCAATCACCCCTCTGTGGGTCC
 GATTTCGAGGAAAGAAGCGGCAGCAACCTGAACAAGTACAGAATCCTGACCGAGCAGCTGCAC
 ACCGAGAAGCTGAAGAAAAAGCTGACAGTGCAGCTGGACCGGCTGATCTACCCCTACAGAATC
 5 TGGCGGCTGGGAAGAGAAGGGCAAAGTGGACATTGTGCTGCTGCCCAGCCGGCAGTTCTACA
 ACCAGATCTTCCTGGACATCGAGGAAAAGGGCAAGCACGCCTTCACCTACAAGGATGAGAGC
 ATCAAGTTCCCTCTGAAGGGCACACTCGGCGGAGCCAGAGTGCAGTTCGACAGAGATCACCT
 GAGAAGATACCCTCACAAGGTGGAAAGCGGCAACGTGGGCAGAATCTACTTCAACATGACCG
 TGAACATCGAGCCTACAGAGTCCCCAGTGTCCAAGTCTCTGAAGATCCACCGGGACGACTTC
 10 CCCAAGGTGGTCAACTTCAAGCCCAAAGAACTGACCGAGTGGATCAAGGACAGCAAGGGCAA
 GAAACTGAAGTCCGGCATCGAGTCCCTGGAAATCGGCCTGAGAGTGATGAGCATCGACCTGG
 GACAGAGACAGGCCGCTGCCGCCTCTATTTTCGAGGTGGTGGATCAGAAGCCCGACATCGAA
 GGCAAGCTGTTTTTCCCAATCAAGGGCACCGAGCTGTATGCCGTGCACAGAGCCAGCTTCAA
 CATCAAGCTGCCCCGGCGAGACACTGGTCAAGAGCAGAGAAGTGCTGCGGAAGGCCAGAGAGG
 15 ACAATCTGAAACTGATGAACCAGAAGCTCAACTTCCTGCGGAACGTGCTGCACTTCCAGCAG
 TTCGAGGACATCACCGAGAGAGAGAAGCGGGTCACCAAGTGGATCAGCAGACAAGAGAACAG
 CGACGTGCCCCTGGTGTACCAGGATGAGCTGATCCAGATCCGCGAGCTGATGTACAAGCCTT
 ACAAGGACTGGGTGCCTTCCTGAAGCAGCTCCACAAGAGACTGGAAGTCGAGATCGGCAAA
 GAAGTGAAGCACTGGCGGAAGTCCCTGAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCCT
 20 GAAGAACATCGACGAGATCGATCGGACCCGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTA
 CCGAACCTGGCGAAGTGCGTAGACTGGAACCCGGCCAGAGATTGCGCCATCGACCAGCTGAAT
 CACCTGAACGCCCTGAAAGAAGATCGGCTGAAGAAGATGGCCAACACCATCATCATGCACGC
 CCTGGGCTACTGCTACGACGTGCGGAAGAAGAAATGGCAGGCTAAGAACCCCGCCTGCCAGA
 TCATCCTGTTCGAGGATCTGAGCAACTACAACCCCTACGAGGAAAGGTCCCGCTTCGAGAAC
 25 AGCAAGCTCATGAAGTGGTCCAGACGCGAGATCCCCAGACAGGTTGCACTGCAGGGCGAGAT
 CTATGGCCTGCAAGTGGGAGAAGTGGGCGCTCAGTTCAGCAGCAGATTCCACGCCAAGACAG
 GCAGCCCTGGCATCAGATGTAGCGTCGTGACCAAAGAGAAGCTGCAGGACAATCGGTTCTTC
 AAGAATCTGCAGAGAGAGGGCAGACTGACCCTGGACAAAATCGCCGTGCTGAAAGAGGGCGA
 TCTGTACCCAGACAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAGGATCGGAAGTGCGTGA
 30 CCACACACGCCGACATCAACGCCGCTCAGAACCTGCAGAAGCGGTTCTGGACAAGAACCCAC
 GGCTTCTACAAGGTGTACTGCAAGGCCTACCAGGTGGACGGCCAGACCGTGTACATCCCT
 GAGAGCAAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGGGCTACTTCATTCTGAA
 GGACGGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAAGAAGGGCAGCTCCAAGC
 AGAGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCTGGCCTCCGAG

CTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCAGCGGCAATGTGTTCCCCAGCGACAA
 ATGGATGGCCGCTGGCGTGTCTTCGGAAAGCTGGAACGCATCCTGATCAGCAAGCTGACCA
 ACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTATGAAAAGGCCGGCG
GCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCC TACCCATACGATGTTCCAGA
 5 TTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCGACTATGCCT
 AA

MAPKKRKVGIIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAI
 YEHHEQDPKNPKKVSKEIQAEIQLWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVE
 10 KKGEANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPGSGSGSSEVEFSHEYWM
 RHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYR
 LYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGI
 LADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSWEEEEKKKWEE
 DKKKDPLAKILGKLAEYGLIPLFIPTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALER
 15 FLSWESWNLKVKEEYKVEKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRL
 SKRGLRGWREIIQKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRN
 HPEYPYLYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTE
 KLKKKLTVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIK
 FPLKGTLGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPK
 20 VVNFKPKELTEWIKDSKGKKLKSIESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIEGK
 LFFPIKGTELYAVHRASFNIKLPGETLVKSREVLARKAREDNLKL MNQKLNFLRNVLHFQQFE
 DITEREKRVTKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFKQLHHRLEVEIGKEV
 KHWKSLSDGRKGLYISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHL
 NALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYPYEERSRFENSK
 25 LMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRRFFKN
 LQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGF
 YKVYCKAYQVDGQTVYIPESKDQKQKIIIEFGEYFILKDGVEYEWV NAGKLKIKKGSSKQSS
 SELVDS DILKDSFDLASELKGEKLMLYRDPGSGNVFSPDKWMAAGVFFGKLERILISKLTNQY
 SISTIEDDSSKQSMKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA

30

BhCas12b GSGGS-ABE8-Xten20 at K255

GCCACCATGGCCCCAAGAAGAAGCGGAAGGTTCGGTATCCACGGAGTCCCAGCAGCCGCCAC
 CAGATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAAACCC
 ACGAGGTGCTGAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGGCAAGAG

GCCATCTACGAGCACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAGGCCGAGAT
 CCAGGCCGAGCTGTGGGATTTCTGTGCTGAAGATGCAGAAGTGCAACAGCTTCACACACGAGG
 TGGACAAGGACGAGGTGTTCAACATCCTGAGAGAGCTGTACGAGGAAGTGGTGCCACGAGC
 GTGGAAAAGAAGGGCGAAGCCAACCAGCTGAGCAACAAGTTTCTGTACCCTCTGGTGGACCC
 5 CAACAGCCAGTCTGGAAAGGGAACAGCCAGCAGCGGCAGAAAGCCCAGATGGTACAACCTGA
 AGATTGCCGGCGATCCCTCCTGGGAAGAAGAGAAGAAGAAGTGGGAAGAAGATAAGAAAAAG
 GACCCGCTGGCCAAGATCCTGGGCAAGCTGGCTGAGTACGGACTGATCCCTCTGTTCATCCC
 CTACACCGACAGCAACGAGCCCATCGTGAAAGAAATCAAGTGGATGGAAAAGTCCCGGAACC
 AGAGCGTGCGGCGGCTGGATAAGGACATGTTCAATTCAGGCCCTGGAACGGTTCCTGAGCTGG
 10 GAGAGCTGGAACCTGAAAGTGAAAGAGGAATACGAGAAGGTGAGAAAGAGTACAAGACCCT
 GGAAGAGAGGATCAAAGgaggctctggaggaagcTCCGAAGTCGAGTTTTTCCCATGAGTACT
GGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGG
GCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCA
CGACCCCACTGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATT
 15 ATCGACTTTATGATGCGACGCTGTACGTACGTTTGAACCTTGCGTAATGTGCGCGGGAGCT
ATGATTCACTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGC
AGGTTCACTGATGGACGTGCTGCATCATCCAGGCATGAACCACCGGTAGAAATCACAGAAG
GCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTCTTTTTTTCGCATGCCAGGCGGGTC
TTTAACGCCCAGAAAAAGCACAAATCCTCTACTGACGGCTCTTCTGGATCTGAAACACCTGG
 20 CACAAGCGAGAGCGCCACCCCTGAGAGCTCTGGCGAGGACATCCAGGCTCTGAAGGCTCTGG
 AACAGTATGAGAAAGAGCGGCAAGAACAGCTGCTGCGGGACACCCTGAACACCAACGAGTAC
 CGGCTGAGCAAGAGAGGCCTTAGAGGCTGGCGGGAAATCATCCAGAAATGGCTGAAAATGGA
 CGAGAACGAGCCCTCCGAGAAGTACCTGGAAGTGTTCAAGGACTACCAGCGGAAGCACCTA
 GAGAGGCCGGCGATTACAGCGTGTACGAGTTCCTGTCCAAGAAAGAGAACCACTTCATCTGG
 25 CGGAATCACCTGAGTACCCCTACCTGTACGCCACCTTCTGCGAGATCGACAAGAAAAAGAA
 GGACGCCAAGCAGCAGGCCACCTTCACACTGGCCGATCCTATCAATCACCTCTGTGGGTCC
 GATTCGAGGAAAGAAGCGGCAGCAACCTGAACAAGTACAGAATCCTGACCGAGCAGCTGCAC
 ACCGAGAAGCTGAAGAAAAAGCTGACAGTGCAGCTGGACCGGCTGATCTACCCTACAGAATC
 TGGCGGCTGGGAAGAGAAGGGCAAAGTGGACATTGTGCTGCTGCCCAGCCGGCAGTTCTACA
 30 ACCAGATCTTCTGGACATCGAGGAAAAGGGCAAGCACGCCTTCACCTACAAGGATGAGAGC
 ATCAAGTTCCTCTGAAGGGCACACTCGGCGGAGCCAGAGTGCAGTTCGACAGAGATCACCT
 GAGAAGATACCCTCACAAGGTGGAAAGCGGCAACGTGGGCAGAATCTACTTCAACATGACCG
 TGAACATCGAGCCTACAGAGTCCCCAGTGTCCAAGTCTCTGAAGATCCACCGGGACGACTTC
 CCCAAGGTGGTCAACTTCAAGCCCAAAGAACTGACCGAGTGGATCAAGGACAGCAAGGGCAA

GAAACTGAAGTCCGGCATCGAGTCCCTGGAAATCGGCCTGAGAGTGATGAGCATCGACCTGG
 GACAGAGACAGGCCGCTGCCGCCTCTATTTTCGAGGTGGTGGATCAGAAGCCCGACATCGAA
 GGCAAGCTGTTTTTCCCAATCAAGGGCACCGAGCTGTATGCCGTGCACAGAGCCAGCTTCAA
 CATCAAGCTGCCCCGGCGAGACACTGGTCAAGAGCAGAGAAGTGCTGCGGAAGGCCAGAGAGG
 5 ACAATCTGAAACTGATGAACCAGAAGCTCAACTTCCTGCGGAACGTGCTGCACTTCCAGCAG
 TTCGAGGACATCACCGAGAGAGAGAAGCGGGTCACCAAGTGGATCAGCAGACAAGAGAACAG
 CGACGTGCCCCTGGTGTACCAGGATGAGCTGATCCAGATCCGCGAGCTGATGTACAAGCCTT
 ACAAGGACTGGGTGCGCTTCCTGAAGCAGCTCCACAAGAGACTGGAAGTCGAGATCGGCAAA
 GAAGTGAAGCACTGGCGGAAGTCCCTGAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCCT
 10 GAAGAACATCGACGAGATCGATCGGACCCGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTA
 CCGAACCTGGCGAAGTGCCTAGACTGGAACCCGGCCAGAGATTGCGCCATCGACCAGCTGAAT
 CACCTGAACGCCCTGAAAGAAGATCGGCTGAAGAAGATGGCCAACACCATCATCATGCACGC
 CCTGGGCTACTGCTACGACGTGCGGAAGAAGAAATGGCAGGCTAAGAACCCCGCCTGCCAGA
 TCATCCTGTTCGAGGATCTGAGCAACTACAACCCCTACGAGGAAAGGTCCCGCTTCGAGAAC
 15 AGCAAGCTCATGAAGTGGTCCAGACGCGAGATCCCCAGACAGGTTGCACTGCAGGGCGAGAT
 CTATGGCCTGCAAGTGGGAGAAGTGGGCGCTCAGTTCAGCAGCAGATTCCACGCCAAGACAG
 GCAGCCCTGGCATCAGATGTAGCGTCGTGACCAAAGAGAAGCTGCAGGACAATCGGTTCTTC
 AAGAATCTGCAGAGAGAGGGCAGACTGACCCTGGACAAAATCGCCGTGCTGAAAGAGGGCGA
 TCTGTACCCAGACAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAGGATCGGAAGTGCGTGA
 20 CCACACACGCCGACATCAACGCCGCTCAGAACCTGCAGAAGCGGTTCTGGACAAGAACCCAC
 GGCTTCTACAAGGTGTACTGCAAGGCCTACCAGGTGGACGGCCAGACCGTGTACATCCCTGA
 GAGCAAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGGGCTACTTCATTCTGAAGG
 ACGGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAAGAAGGGCAGCTCCAAGCAG
 AGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCTGGCCTCCGAGCT
 25 GAAAGGCGAAAAGCTGATGCTGTACAGGGACCCAGCGGCAATGTGTTCCCCAGCGACAAAT
 GGATGGCCGCTGGCGTGTCTTCGAAAGCTGGAACGCATCCTGATCAGCAAGCTGACCAAC
 CAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTATGAAAAGGCCGGCGGC
CACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCC
.....TACCCATACGATGTTCCAGATT
ACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCGACTATGCCTAA
 30 MAPKKKRKVGIIHGVPAATRSLFILKIEPNEEVKKGLWKTHEVLNHGIIAYYMNILKLIRQEAI
 YEHHEQDPKNPKKVSKEIQAEIQLWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVE
 KKGEANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEKKKKWEEDKKKDP
 LAKILGKLAEYGLIPLFIPTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWES

WNLKVKEEYKVEKEYKTLEERIKGGSSGSSEVEFSHEYWMRHALTLAKRARDEREVPVGAV
 LVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMI
 HSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFN
 AQKKAQSSTDGSSGSETPGTSESATPESSGEDIQALKALEQYEKERQEQLLRDTLNTNEYRL
 5 SKRGLRGWREIIQKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRN
 HPEYPYLYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTE
 KLKKKLTVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIK
 FPLKGTGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDDFPK
 VVNFKPKELTEWIKDSKGKKLKS GIESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIEGK
 10 LFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRLKAREDNLKLMNQKLNFLRNLVLFHQQFE
 DITEREKRVTKWISRQENS DVPLVYQDELIQIRELMYKPYKDWVAFKQLHKRLEVEIGKEV
 KHWKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFQFAIDQLNHL
 NALKEDRLKKMANTIIMHALGYCYDVRKKKKWQAKNPACQIILFEDLSNYPYEERSRFENSK
 LMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPIRCSVVTKEKLQDNRRFFKN
 15 LQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGF
 YKVYCKAYQVDGQTVYIPESKDQKQKIIIEFGEYFILKDGVEYEWVWAGKLKIKKGSSKQSS
 SELVDS DILKDSFDLASELKGEKLMLYRDP SGNVFPSDKWMAAGVFFGKLERILISKLTNQY
 SISTIEDSSKQSMKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA

20 BhCas12b GGS GGS-ABE8-Xten20 at D306

GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTTCGGTATCCACGGAGTCCCAGCAGCCGCCAC
 CAGATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAAACCC
 ACGAGGTGCTGAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGGCAAGAG
 GCCATCTACGAGCACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAGGCCGAGAT
 25 CCAGGCCGAGCTGTGGGATTTCTGTGCTGAAGATGCAGAAGTGCAACAGCTTCACACACGAGG
 TGGACAAGGACGAGGTGTTCAACATCCTGAGAGAGCTGTACGAGGAAGTGGTGCCACGAGC
 GTGGAAAAGAAGGGCGAAGCCAACCAGCTGAGCAACAAGTTTCTGTACCCTCTGGTGGACCC
 CAACAGCCAGTCTGGAAAGGGAACAGCCAGCAGCGGCAGAAAGCCCAGATGGTACAACCTGA
 AGATTGCCGGCGATCCCTCCTGGGAAGAAGAGAAGAAGAAGTGGGAAGAAGATAAGAAAAAG
 30 GACCCGCTGGCCAAGATCCTGGGCAAGCTGGCTGAGTACGGAAGTATCCCTCTGTTCATCCC
 CTACACCGACAGCAACGAGCCCATCGTGAAAGAAATCAAGTGGATGGAAAAGTCCCGGAACC
 AGAGCGTGCGGCGGCTGGATAAGGACATGTTTCATTTCAGGCCCTGGAACGGTTCCTGAGCTGG
 GAGAGCTGGAACCTGAAAGTGAAGAGGAATACGAGAAGGTGAGAAAGAGTACAAGACCCT
 GGAAGAGAGGATCAAAGAGGACATCCAGGCTCTGAAGGCTCTGGAACAGTATGAGAAAGAGC

GGCAAGAACAGCTGCTGCGGGACACCCTGAACACCAACGAGTACCGGCTGAGCAAGAGAGGC
 CTTAGAGGCTGGCGGGAAATCATCCAGAAATGGCTGAAAATGGACGgaggctctggagggaag
 cTCCGAAGTTCGAGTTTTCCTCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGG
 CTCGAGATGAACGCGAGGTGCCCCTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATCGGC
 5 GAAGGTTTGAATAGGGCAATCGGACTCCACGACCCCACTGCACATGCGGAAATCATGGCCCT
 TCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTTATGATGCGACGCTGTACGTCACGT
 TTGAACCTTGCCTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTTGGACGAGTTGTATTC
 GGTGTTTCGCAACGCCAAGACGGGTGCCGAGGTTCACTGATGGACGTGCTGCATCATCCAGG
 CATGAACCACCGGGTAGAAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGT
 10 GTCGTTTTTTTTTCGCATGCCAGGCGGGTCTTTAACGCCAGAAAAAGCACAAATCCTCTACT
GACGGCTCTTCTGGATCTGAAACACCTGGCACAAGCGAGAGCGCCACCCCTGAGAGCTCTGG
CGAGAACGAGCCCTCCGAGAAGTACCTGGAAGTGTTCAAGGACTACCAGCGGAAGCACCCCTA
 GAGAGGCCGGCGATTACAGCGTGTACGAGTTCCTGTCCAAGAAAGAGAACCACCTTCATCTGG
 CGGAATCACCCCTGAGTACCCCTACCTGTACGCCACCTTCTGCGAGATCGACAAGAAAAAGAA
 15 GGACGCCAAGCAGCAGGCCACCTTCACACTGGCCGATCCTATCAATCACCCCTCTGTGGGTCC
 GATTTCGAGGAAAGAAGCGGCAGCAACCTGAACAAGTACAGAATCCTGACCGAGCAGCTGCAC
 ACCGAGAAGCTGAAGAAAAAGCTGACAGTGCAGCTGGACCGGCTGATCTACCCTACAGAATC
 TGGCGGCTGGGAAGAGAAGGGCAAAGTGGACATTGTGCTGCTGCCAGCCGGCAGTTCTACA
 ACCAGATCTTCTGGACATCGAGGAAAAGGGCAAGCACGCCTTCACCTACAAGGATGAGAGC
 20 ATCAAGTTCCTCTGAAGGGCACACTCGGCGGAGCCAGAGTGCAGTTCGACAGAGATCACCT
 GAGAAGATACCCTCACAAGGTGGAAAGCGGCAACGTGGGCAGAATCTACTTCAACATGACCG
 TGAACATCGAGCCTACAGAGTCCCCAGTGTCCAAGTCTCTGAAGATCCACCGGGACGACTTC
 CCCAAGGTGGTCAACTTCAAGCCCAAAGAACTGACCGAGTGGATCAAGGACAGCAAGGGCAA
 GAACTGAAGTCCGGCATCGAGTCCCTGGAAATCGGCCTGAGAGTGATGAGCATCGACCTGG
 25 GACAGAGACAGGCCGCTGCCGCTCTATTTTCGAGGTGGTGGATCAGAAGCCCGACATCGAA
 GGCAAGCTGTTTTTCCCAATCAAGGGCACCGAGCTGTATGCCGTGCACAGAGCCAGCTTCAA
 CATCAAGCTGCCCCGGCGAGACACTGGTCAAGAGCAGAGAAGTGCTGCGGAAGGCCAGAGAGG
 ACAATCTGAACTGATGAACCAGAAGCTCAACTTCCTGCGGAACGTGCTGCACTTCCAGCAG
 TTCGAGGACATCACCGAGAGAGAGAAGCGGGTCACCAAGTGGATCAGCAGACAAGAGAACAG
 30 CGACGTGCCCTGGTGTACCAGGATGAGCTGATCCAGATCCGCGAGCTGATGTACAAGCCTT
 ACAAGGACTGGGTGCCTTCCTGAAGCAGCTCCACAAGAGACTGGAAGTCGAGATCGGCCAAA
 GAAGTGAAGCACTGGCGGAAGTCCCTGAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCCT
 GAAGAACATCGACGAGATCGATCGGACCCGGAAGTTCTTGCTGAGATGGTCCCTGAGGCCTA
 CCGAACCTGGCGAAGTGCCTAGACTGGAACCCGGCCAGAGATTGCCATCGACCAGCTGAAT

CACCTGAACGCCCTGAAAGAAGATCGGCTGAAGAAGATGGCCAACACCATCATCATGCACGC
 CCTGGGCTACTGCTACGACGTGCGGAAGAAGAAATGGCAGGCTAAGAACCCCGCCTGCCAGA
 TCATCCTGTTCGAGGATCTGAGCAACTACAACCCCTACGAGGAAAGGTCCCGCTTCGAGAAC
 AGCAAGCTCATGAAGTGGTCCAGACGCGAGATCCCCAGACAGGTTGCACTGCAGGGCGAGAT
 5 CTATGGCCTGCAAGTGGGAGAAGTGGGCGCTCAGTTCAGCAGCAGATTCCACGCCAAGACAG
 GCAGCCCTGGCATCAGATGTAGCGTCGTGACCAAAGAGAAGCTGCAGGACAATCGGTTCTTC
 AAGAATCTGCAGAGAGAGGGCAGACTGACCCTGGACAAAATCGCCGTGCTGAAAGAGGGCGA
 TCTGTACCCAGACAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAGGATCGGAAGTGCCTGA
 CCACACACGCCGACATCAACGCCGCTCAGAACCTGCAGAAGCGGTTCTGGACAAGAACCCAC
 10 GGCTTCTACAAGGTGTACTGCAAGGCCTACCAGGTGGACGGCCAGACCGTGTACATCCCTGA
 GAGCAAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGGGCTACTTCATTCTGAAGG
 ACGGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAAGAAGGGCAGCTCCAAGCAG
 AGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCTGGCCTCCGAGCT
 GAAAGGCGAAAAGCTGATGCTGTACAGGGACCCAGCGGCAATGTGTTCCCCAGCGACAAAT
 15 GGATGGCCGCTGGCGTGTTCTTCGAAAGCTGGAACGCATCCTGATCAGCAAGCTGACCAAC
 CAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTATGAAAAGGCCGGCGGC
CACGAAAAAGGCCGGCCAGGCCAAAAAAGAAAAAGGGATCCTACCCATACGATGTTCCAGATT
ACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCGACTATGCCTAA

20 MAPKKKRKVGIIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAI
 YEHHEQDPKNPKKVSKEIQAELWDFVLKMQKNSFTHEVDKDEVFNILRELYEELVPSSVE
 KKGEANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEKKKWEEDKKKDP
 LAKILGKLAEYGLIPLFIPTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWES
 WNLKVKEEYKVEKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLR
 25 GWREIIQKWLKMDGGSGGSSEVEFSHEYWMRHALLAKRARDEREVPVGAVLVLNNRVIGEG
 WNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGV
 RNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDG
 SSGSETPGTSESATPESSGENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRN
 HPEYPYLYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTE
 30 KLKKKLTVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIK
 FPLKGTLGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIPTESPVSKSLKIHRDDFPK
 VVNFKPKELTEWIKDSKGKKLKSIESLEIGLRVMSIDLQQRQAAAASIFEVVDQKPDIEGK
 LFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRLKAREDNLKLMNQKLNFLRNVLHFQQFE
 DITEREKRVTKWISRQENS DVPLVYQDELIQIRELMYKPYKDWVAFKQLHKRLEVEIGKEV

KHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHL
 NALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYPYEERSRFENSK
 LMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRRFFKN
 LQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGF
 5 YKVYCKAYQVDGQTVYIPESKDQKQKIIIEFGEGYFILKDGVEWVNAGKLKIKGSSKQSS
 SELVDSILKDSFDLASELKGEKLMLYRDPGNGVFPSPDKWMAAGVFFGKLERILISKLTNQY
 SISTIEDSSKQSMKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA

BhCas12b GSGGS-ABE8-Xten20 at D980

10 GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTTCGGTATCCACGGAGTCCCAGCAGCCGCCAC
 CAGATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAAACCC
 ACGAGGTGCTGAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGGCAAGAG
 GCCATCTACGAGCACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAGGCCGAGAT
 CCAGGCCGAGCTGTGGGATTTCTGTGCTGAAGATGCAGAAGTGCAACAGCTTCACACACGAGG
 15 TGGACAAGGACGAGGTGTTCAACATCCTGAGAGAGCTGTACGAGGAAGTGGTGCCCAGCAGC
 GTGGAAAAGAAGGGCGAAGCCAACAGCTGAGCAACAAGTTTCTGTACCCTCTGGTGAGCCC
 CAACAGCCAGTCTGGAAAGGGAACAGCCAGCAGCGGCAGAAAGCCCAGATGGTACAACCTGA
 AGATTGCCGGCGATCCCTCCTGGGAAGAAGAGAAGAAGAAGTGGGAAGAAGATAAGAAAAAG
 GACCCGCTGGCCAAGATCCTGGGCAAGCTGGCTGAGTACGGAAGTATCCCTCTGTTCATCCC
 20 CTACACCGACAGCAACGAGCCCATCGTGAAAGAAATCAAGTGGATGGAAAAGTCCCAGAAC
 AGAGCGTGCGGGGCTGGATAAGGACATGTTTCATTACAGGCCCTGGAACGGTTCCTGAGCTGG
 GAGAGCTGGAACCTGAAAGTGAAGAGGAATACGAGAAGGTGAGAAAGAGTACAAGACCCT
 GGAAGAGAGGATCAAAGAGGACATCCAGGCTCTGAAGGCTCTGGAACAGTATGAGAAAGAGC
 GGCAAGAACAGCTGCTGCGGGACACCCTGAACACCAACGAGTACCGGCTGAGCAAGAGAGGC
 25 CTTAGAGGCTGGCGGGAAATCATCCAGAAATGGCTGAAAATGGACGAGAACGAGCCCTCCGA
 GAAGTACCTGGAAGTGTTCAGGACTACCAGCGGAAGCACCTAGAGAGGCCGGCGATTACA
 GCGTGTACGAGTTCCTGTCCAAGAAAGAGAACCACTTCATCTGGCGGAATCACCTGAGTAC
 CCCTACCTGTACGCCACCTTCTGCGAGATCGACAAGAAAAAGAAGGACGCCAAGCAGCAGGC
 CACCTTCACACTGGCCGATCCTATCAATCACCTCTGTGGGTCCGATTGAGGAAAGAAGCG
 30 GCAGCAACCTGAACAAGTACAGAATCCTGACCGAGCAGCTGCACACCGAGAAGCTGAAGAAA
 AAGCTGACAGTGCAGCTGGACCGGCTGATCTACCCTACAGAATCTGGCGGCTGGGAAGAGAA
 GGGCAAAGTGGACATTGTGCTGCTGCCCAGCCGGCAGTTCTACAACCAGATCTTCCTGGACA
 TCGAGGAAAAGGGCAAGCACGCCTTCACCTACAAGGATGAGAGCATCAAGTTCCTCTGAAG
 GGCACACTCGGCGGAGCCAGAGTGCAGTTCGACAGAGATCACCTGAGAAGATACCCTCACAA

GGTGGAAAGCGGCAACGTGGGCAGAATCTACTTCAACATGACCGTGAACATCGAGCCTACAG
 AGTCCCCAGTGTCCAAGTCTCTGAAGATCCACCGGGACGACTTCCCCAAGGTGGTCAACTTC
 AAGCCCAAAGAACTGACCGAGTGGATCAAGGACAGCAAGGGCAAGAACTGAAGTCCGGCAT
 CGAGTCCCTGGAAATCGGCCTGAGAGTGATGAGCATCGACCTGGGACAGAGACAGGCCGCTG
 5 CCGCCTCTATTTTCGAGGTGGTGGATCAGAAGCCCGACATCGAAGGCAAGCTGTTTTTCCCA
 ATCAAGGGCACCGAGCTGTATGCCGTGCACAGAGCCAGCTTCAACATCAAGCTGCCCCGGCGA
 GAACTGGTCAAGAGCAGAGAAGTGCTGCGGAAGGCCAGAGAGGACAATCTGAACTGATGA
 ACCAGAAGCTCAACTTCCTGCGGAACGTGCTGCACTTCCAGCAGTTCGAGGACATCACCGAG
 AGAGAGAAGCGGGTCACCAAGTGGATCAGCAGACAAGAGAACAGCGACGTGCCCCTGGTGTA
 10 CCAGGATGAGCTGATCCAGATCCGCGAGCTGATGTACAAGCCTTACAAGGACTGGGTGCGCT
 TCCTGAAGCAGCTCCACAAGAGACTGGAAGTCGAGATCGGCAAAGAAGTGAAGCACTGGCGG
 AAGTCCCTGAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAGAACATCGACGAGAT
 CGATCGGACCCGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACCGAACCTGGCGAAGTGC
 GTAGACTGGAACCCGGCCAGAGATTGCGCATCGACCAGCTGAATCACCTGAACGCCCTGAAA
 15 GAAGATCGGCTGAAGAAGATGGCCAACACCATCATCATGCACGCCCTGGGCTACTGCTACGA
 CGTGCGGAAGAAGAAATGGCAGGCTAAGAACCCCGCCTGCCAGATCATCCTGTTCGAGGATC
 TGAGCAACTACAACCCCTACGAGGAAAGGTCCCGCTTCGAGAACAGCAAGCTCATGAAGTGG
 TCCAGACGCGAGATCCCCAGACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTGGG
 AGAAGTGGGCGCTCAGTTCAGCAGCAGATTCCACGCCAAGACAGGCAGCCCTGGCATCAGAT
 20 GTAGCGTCGTGACCAAAGAGAAGCTGCAGGACAATCGGTTCTTCAAGAATCTGCAGAGAGAG
 GGCAGACTGACCCTGGACAAAATCGCCGTGCTGAAAGAGGGCGATCTGTACCCAGACAAAGG
 CGGCGAGAAGTTCATCAGCCTGAGCAAGGATCGGAAGTGCGTGACCACACACGCCGACATCA
 ACGCCGCTCAGAACCTGCAGAAGCGGTTCTGGACAAGAACCCACGGCTTCTACAAGGTGTAC
 TGCAAGGCCTACCAGGTGGACgagagctctggaggaagcTCCGAAGTCGAGTTTTCCCATGA
 25 GTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCCG
 TGGGGGCGAGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGA
 CTCCACGACCCCACTGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCA
 GAATTATCGACTTTATGATGCGACGCTGTACGTACGTTTGAACCTTGCGTAATGTGCGCGG
 GAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTGGTGTTTCGCAACGCCAAGACGGGT
 30 GCCGCGAGGTTCACTGATGGACGTGCTGCATCATCCAGGCATGAACCACCGGGTAGAAATCAC
 AGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTCGTTTTTTTCGCATGCCAGGC
 GGGTCTTTAACGCCCAAGAAAAAGCACAATCCTCTACTGACGGCTCTTCTGGATCTGAAACA
 CCTGGCACAAGCGAGAGCGCCACCCCTGAGAGCTCTGGCGGCCAGACCGTGTACATCCCTGA
 GAGCAAGGACCAGAAGCAGAAGATCATCGAAGAGTTTCGGCGAGGGCTACTTCATTCTGAAGG

ACGGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAAGAAGGGCAGCTCCAAGCAG
 AGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCTGGCCTCCGAGCT
 GAAAGGCGAAAAGCTGATGCTGTACAGGGACCCAGCGGCAATGTGTTCCCCAGCGACAAAT
 GGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCCTGATCAGCAAGCTGACCAAC
 5 CAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTATGAAAAGGCCGGCGGC
CACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCCTACCCATACGATGTTCCAGATT
ACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCGACTATGCCTAA

 MAPKKKRKVGIIHGVPAATRSLFKIEPNEEVKKGLWKTHEVLNHGIIAYYMNILKLIRQEAI
 10 YEHHEQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVE
 KKGEANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEKKKKWEEDKKKDP
 LAKILGKLAEYGLIPLFIPTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWES
 WNLKVKEEYKVEKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLR
 GWREIIQKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPY
 15 LYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKL
 TVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKKGKHAFTYKDESIKFPKGT
 LGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKP
 KELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLQGRQAAAASIFEVVDQKPDIEGKLLFFPIK
 GTELYAVHRASFNIKLPGETLVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDITERE
 20 KRVTKWISRQENS DVPLVYQDELIQIRELMYKPYKDWVAFKQLHHRLEVEIGKEVKHWRKS
 LSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFADQLNHLNALKED
 RLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYPYEERSRFENSKLMKWSR
 REIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFKNLQREGR
 LTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCK
 25 AYQVDGGSGGSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLH
 DPTAHAEIMALRQGGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAA
 GSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPG
 TSESATPESGGQT VYIPE SKDQKQKII EEFGEYFILKDG VYEWVNAGKLKIKKGSSKQSS
 SELVDS DILKDSFDLASELKGEKLMLYRDP SGNVFP SDKWMAAGVFFGKLERILISKLTNQY
 30 SISTIEDSSKQSMKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA

BhCas12b GSGGS-ABE8-Xten20 at K1019

GCCACCATGGCCCCAAAGAAGAAAGCGGAAGGTCCGTATCCACGGAGTCCCAGCAGCCGCCAC
 CAGATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAAACCC

ACGAGGTGCTGAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGGCAAGAG
 GCCATCTACGAGCACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAGGCCGAGAT
 CCAGGCCGAGCTGTGGGATTTCTGTGCTGAAGATGCAGAAGTGCAACAGCTTCACACACGAGG
 TGGACAAGGACGAGGTGTTCAACATCCTGAGAGAGCTGTACGAGGAACTGGTGCCCAGCAGC
 5 GTGGAAAAGAAGGGCGAAGCCAACCAGCTGAGCAACAAGTTTCTGTACCCTCTGGTGGACCC
 CAACAGCCAGTCTGGAAAGGGAACAGCCAGCAGCGGCAGAAAGCCCAGATGGTACAACCTGA
 AGATTGCCGGCGATCCCTCCTGGGAAGAAGAGAAGAAGAAGTGGGAAGAAGATAAGAAAAAG
 GACCCGCTGGCCAAGATCCTGGGCAAGCTGGCTGAGTACGGACTGATCCCTCTGTTCATCCC
 CTACACCGACAGCAACGAGCCCATCGTGAAAGAAATCAAGTGGATGGAAAAGTCCCGGAACC
 10 AGAGCGTGCGGCGGCTGGATAAGGACATGTTCAATTCAGGCCCTGGAACGGTTCCTGAGCTGG
 GAGAGCTGGAACCTGAAAGTGAAAGAGGAATACGAGAAGGTGAGAAAGAGTACAAGACCCT
 GGAAGAGAGGATCAAAGAGGACATCCAGGCTCTGAAGGCTCTGGAACAGTATGAGAAAGAGC
 GGCAAGAACAGCTGCTGCGGGACACCCTGAACACCAACGAGTACCGGCTGAGCAAGAGAGGC
 CTTAGAGGCTGGCGGGAAATCATCCAGAAATGGCTGAAAATGGACGAGAACGAGCCCTCCGA
 15 GAAGTACCTGGAAGTGTTCAAGGACTACCAGCGGAAGCACCCCTAGAGAGGCCGGCGATTACA
 GCGTGTACGAGTTCTGTCCAAGAAAGAGAACCCTTCATCTGGCGGAATCACCTGAGTAC
 CCCTACCTGTACGCCACCTTCTGCGAGATCGACAAGAAAAAGAAGGACGCCAAGCAGCAGGC
 CACCTTCACACTGGCCGATCCTATCAATCACCTCTGTGGGTCCGATTGAGGAAAGAAGCG
 GCAGCAACCTGAACAAGTACAGAATCCTGACCGAGCAGCTGCACACCGAGAAGCTGAAGAAA
 20 AAGCTGACAGTGACAGCTGGACCGGCTGATCTACCCTACAGAATCTGGCGGCTGGGAAGAGAA
 GGGCAAAGTGACATTGTGCTGCTGCCAGCCGGCAGTTCTACAACCAGATCTTCCTGGACA
 TCGAGGAAAAGGGCAAGCACGCCTTCACCTACAAGGATGAGAGCATCAAGTTCCTCTGAAG
 GGCACACTCGGCGGAGCCAGAGTGACAGTTCGACAGAGATCACCTGAGAAGATACCCTCACAA
 GGTGGAAAGCGGCAACGTGGGCAGAATCTACTTCAACATGACCGTGAACATCGAGCCTACAG
 25 AGTCCCCAGTGTCCAAGTCTCTGAAGATCCACCGGGACGACTTCCCCAAGGTGGTCAACTTC
 AAGCCCAAAGAACTGACCGAGTGGATCAAGGACAGCAAGGGCAAGAACTGAAGTCCGGCAT
 CGAGTCCCTGGAAATCGGCCTGAGAGTGATGAGCATCGACCTGGGACAGAGACAGGCCGCTG
 CCGCCTCTATTTTCGAGGTGGTGGATCAGAAGCCCGACATCGAAGGCAAGCTGTTTTTCCCA
 ATCAAGGGCACCGAGCTGTATGCCGTGCACAGAGCCAGCTTCAACATCAAGCTGCCCCGGCGA
 30 GACACTGGTCAAGAGCAGAGAAGTGCTGCGGAAGGCCAGAGAGGACAATCTGAAACTGATGA
 ACCAGAAGCTCAACTTCCTGCGGAACGTGCTGCACTTCCAGCAGTTCGAGGACATCACCGAG
 AGAGAGAAGCGGGTCACCAAGTGGATCAGCAGACAAGAGAACAGCGACGTGCCCTGGTGTA
 CCAGGATGAGCTGATCCAGATCCGCGAGCTGATGTACAAGCCTTACAAGGACTGGGTGCCT
 TCCTGAAGCAGCTCCACAAGAGACTGGAAGTCGAGATCGGCAAAGAAGTGAAGCACTGGCGG

AAGTCCCTGAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAGAACATCGACGAGAT
 CGATCGGACCCGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACCGAACCTGGCGAAGTGC
 GTAGACTGGAACCCGGCCAGAGATTGCCATCGACCAGCTGAATCACCTGAACGCCCTGAAA
 GAAGATCGGCTGAAGAAGATGGCCAACACCATCATCATGCACGCCCTGGGCTACTGCTACGA
 5 CGTGCGGAAGAAGAAATGGCAGGCTAAGAACCCCGCCTGCCAGATCATCCTGTTCGAGGATC
 TGAGCAACTACAACCCCTACGAGGAAAGGTCCCCTTCGAGAACAGCAAGCTCATGAAGTGG
 TCCAGACGCGAGATCCCCAGACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTGGG
 AGAAGTGGGCGCTCAGTTCAGCAGCAGATTCCACGCCAAGACAGGCAGCCCTGGCATCAGAT
 GTAGCGTCGTGACCAAAGAGAAGCTGCAGGACAATCGGTTCTTCAAGAATCTGCAGAGAGAG
 10 GGCAGACTGACCCTGGACAAAATCGCCGTGCTGAAAGAGGGCGATCTGTACCCAGACAAAGG
 CGGCGAGAAGTTCATCAGCCTGAGCAAGGATCGGAAGTGGTGACACACACGCCGACATCA
 ACGCCGCTCAGAACCTGCAGAAGCGGTTCTGGACAAGAACCACGGCTTCTACAAGGTGTAC
 TGCAAGGCCCTACCAGGTGGACGGCCAGACCGTGTACATCCCTGAGAGCAAGGACCAGAAGCA
 GAAGATCATCGAAGAGTTCGGCGAGGGCTACTTCATTCTGAAGGACGGGGTGTACGAATGGG
 15 TCAACGCCGGCAAGggaggctctggagggaagcTCCGAAGTCGAGTTTTCCCATGAGTACTGG
ATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCGTGGGGGC
AGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACG
ACCCCACTGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTAT
CGACTTTATGATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTAT
 20 GATTCACTCCCGCATTGGACGAGTTGTATTGCGTGTTCGCAACGCCAAGACGGGTGCCGCAG
GTTCACTGATGGACGTGCTGCATCATCCAGGCATGAACCACCGGGTAGAAATCACAGAAGGC
ATATTGGCGGACGAATGTGCGGCGCTGTTGTGTGTTTTTTTCGCATGCCCAGGCGGGTCTT
TAACGCCCAGAAAAAGCACAAATCCTCTACTGACGGCTCTTCTGGATCTGAAACACCTGGCA
CAAGCGAGAGCGCCACCCCTGAGAGCTCTGGCCTGAAAATCAAGAAGGGCAGCTCCAAGCAG
 25 AGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCTGGCCTCCGAGCT
 GAAAGGCGAAAAGCTGATGCTGTACAGGGACCCAGCGGCAATGTGTTCCCCAGCGACAAAT
 GGATGGCCGCTGGCGTGTCTTCGAAAGCTGGAACGCATCCTGATCAGCAAGCTGACCAAC
 CAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTATGAAAAGGCCGGCGGC
CACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCCTACCCATACGATGTTCCAGATT
 30 ACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCCTAA

MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAI
 YEHHEQDPKNPKKVSKEIQAEIQLWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVE
 KKGEANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEKKKKWEEDKKKDP

LAKILGKLAEYGLIPLFIPTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWES
 WNLKVKEEYKVEKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLR
 GWREIIQKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPY
 LYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRIILTEQLHTEKLKKKL
 5 TVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGT
 LGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKP
 KELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLQORQAAAASIFEVVDQKPDIEGKLFFPIK
 GTELYAVHRASFNIKLPGETLVKSREVLARKAREDNLKLMNQKLNFLRNVLHFQQFEDITERE
 KRVTKWISRQENS DVPLVYQDELIQIRELMYKPYKDWVAFLLKQLHQRLEVEIGKEVKHWRKS
 10 LSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKED
 RLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYPYEERSRFENSKLMKWSR
 REIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFKNLQREGR
 LTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCK
 AYQVDGQTVYIPESKDQKQKIIIEFGEGYFILKDGVEWVNAGKGGSGGSSEVEFSHEYWMR
 15 HALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRL
 YDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGIL
 ADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGLKIKKGSSKQSS
 SELVSDILKDSFDLASELKGEKLMLYRDPGSGNVFSPDKWMAAGVFFGKLERILISKLTNQY
 SISTIEDSSKQSMKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA

20 For the sequences above, the Kozak sequence is bolded and underlined;
underline dot underline marks the N-terminal nuclear localization signal (NLS) following the
 Kozak sequence; lower case characters denote the GGGSGGS linker; — — — — marks the
 sequence encoding ABE8, unmodified sequence encodes BhCas12b; double underling
 denotes the Xten20 linker; single underlining denotes the C-terminal NLS; GGATCC denotes
 25 the GS linker; and italicized characters represent the coding sequence of the 3x
 hemagglutinin (HA) tag.

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
 30 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.

5 Exemplary Cas12j/CasΦ amino acid sequences follow:

>CasΦ-1

MADTPTLFTQFLRHHLPGQRFRKDILKQAGRILANKGEDATIAFLRGKSEESPPDFQPPVKC
 PIIACSRPLTEWPIYQASVAIQGYVYGQSLAEFEASDPGCSKDGLLGWFDKTGVCTDYFSVQ
 GLNLI FQNARKRYIGVQTKVTNRNEKRHKKLKRINAKRIAEGLPELTSDEPESALDETGHLI
 10 DPPGLNTNIYCYQQVSPKPLALSEVNQLPTAYAGYSTSGDDPIQPMVTKDRLSISKGPGYI
 PEHQALLSQQKHKRRMRGYGLKARALLVIVRIQDDWAVIDLRSLLRNAYWRRIVQTKEPSTI
 TKLLKLVTGDPVL DATRMVATFTYKPGIVQVRSACKLKNKQGSKLFSERYLNETVSVTSIDL
 GSNNLVAVATYRLVNGNTPPELLQRFTLPSHLVKDFERYKQAHDTLEDSIQKTAVASLPQGQQ
 TEIRMWSMYGFREAQERVQELGLADGSI PWNVMTATSTILTDLFLARGGDPKKCMFTSEPK
 15 KKKNSKQVLYKIRDRAWAKMYRTLLSKETREAWN KALWGLKRGSPDYARLSKRKEELARRCV
 NYTISTA EKRAQCGRTIVALEDLNIGFFHGRGKQEPGWVGLFTRKKENRWLMQALHKA FLEL
 AHHRGYHVIEVNPAYTSQTCPVCRHCDPDNRDQHNREAFHCIGCGFRGNADLDVATHNIAMV
 AITGESLKRARGSVASKTPQPLAAE*

>CasΦ-2

20 MPKPAVESEFSKVLKKHFPGERFRSSYMKRGGKILAAQGEEAVVAYLQ GKSEEEPPNFQPPA
 KCHVVTKS R DFAEWPIMKASEAIQRYIYALSTTERAACKPGKSSESHAAWFAATGVSNHGYS
 HVQGLNLI FDHTLGRYDGV LKKVQLRNEKARARLESINASRADEGLPEIKAE EEEVATNETG
 HLLQPPGINPSFYVYQTISPQAYRPRDEIVLPPEYAGYVRDPNAPIPLGVVRNRCDIQKGCP
 GYIPEWQREAGTAISPKTGKAVTVPGLS PKKNKRMRRYWRSEKEKAQDALLVTVRIGTDWV
 25 IDVRGLLRNARWRTIAPKDISLNALLDLFTGDPVIDVRRNIVTFTYTLDACGTYARKWTLKG
 KQTKATLDKLTATQTVALVAIDL GQTNPI SAGISRVTQENGALQCEPLDRFTLPDDLKDIS
 AYRIAWDRNEEELRARSVEALPEAQQA EVRALDGVSKETARTQLCADFGLDPKRLPWDKMSS
 NTTFISEALLSNSVSRDQVFFTPAPKKGAKKKAPVEVMRKDRTWARAYKPRLSVEAQKLKNE
 ALWALKRTSPEY LKLSRRKEELCRRSINYVIEKTRRRTQCQIVIPVIEDLNVRFFHGS GKRL
 30 PGWDNFFTAKKENRWF IQGLHKA FSDLRTHRSFYVFEVRPERTSITCPKCGHCEVGNRDGEA
 FQCLSCGKTCNADLDVATHNLTQVALTGKTM PKREEPRDAQGTAPARKTKKASKSKAPPAER
 EDQTPAQEPSQTS

>CasΦ-3

MEKEITELTKIRREFPNKKFSSTDMKKAGKLLKAEGPDAVRDFLNSCQEIIIGDFKPPVKTN
 VSISRPFEEWPVSMVGRAIQEYYFSLTKEELESVHPGTSSDHKSFFNITGLSNYNTSVQG
 LNLIFKNAKAIYDGTLVKANNKNKKLEKKFNEINHKSLEGLPIITPDFEFPDENGHLNNP
 PGINRNIYGYQGCAAKVFVPSKHKMVSLPKEYEGYNRDPNLSLAGFRNRLEIPEGEPGHVPW
 5 FQRMDIPEGQIGHVNKIQRFNFVHGKNSGKVKFSDKTGRVKRYHHSKYKDATKPYKFLEESK
 KVSALDSILAIITIGDDWVVFDIRGLYRNVFYRELAQKGLTAVQLLDLFTGDPVIDPKKGVV
 TFSYKEGVVPVFSQKIVPRFKSRDTLEKLTSQGPVALLSVDLGQNEPVAARVCSLKNINDKI
 TLDNSCRISFLDDYKKQIKDYRDSLDELEIKIRLEAINSLETNQQVEIRDLDVFSADRAKAN
 TVDMFDIDPNLISWDSMSDARVSTQISDLYLKNGGDESRVYFEINNKRIRKSDYNISQLVRP
 10 KLS DSTRK NLNDSIWKLKRTSEEYKLKSKRKLELSRAVVNYTIRQSKLLSGINDIVIILEDL
 DVKKKFNGRGIRDIGWDNFFSSRKENRWFIPAFHKA FSELSSNRGLCVIEVNPATWSATCPD
 CGFCSKENRDGINFTCRKCGVSYHADIDVATLNIARVAVLGKPMSPADRERLGDTKKPRVA
 RSRKTMKRKDISNSTVEAMVTA*

>CasΦ-4

15 MY SLEMADLKSEPSLLAKLLRDRFPGKYWLPKYWKLAEEKRLTGGEAAACEYMADKQLDSP
 PNFRPPARCVILAKSRPFEDWPVHRVASKAQSFVIGLSEQGFAALRAAPPSTADARRDWLRS
 HGASEDDLMALEAQLLETIMGNAISLHGGVLKKIDNANVKAARKLSGRNEARLNKGLQELPP
 EQEGSAYGADGLLVNPPGLNLNIYCRKSCCPKPVKNTARFVGHPGYLRDSISILISGTMDR
 LTIIIEGMPGHIPAWQREQGLVKPGRRRRLSGSESNMQRKVDPSTGPRRSTRSGTVNRSNQR
 20 TGRNGDPLLVEIRMKEDWVLLDARGLLRNLRWRESKRGLSCDHEDLSLSGLLALFSGDPVID
 PVRNEVVFLYEGEIIIPVRSTKPVGTRQSKKLLERQASMGPLTLISCDLGQTNLIAGRASAI
 LTHGSLGVRSSVRIELDPEIIKSFERLRKDADRLETEILTAAKETLSDEQRGEVNSHEKDSP
 QTAKASLCRELGLHPPSLPWGQMGPSTTFIADMLISHGRDDDAFLSHGEFPTLEKRKKFKDR
 FCLESRPLLSSETRKALNESLWEVKRTSSEYARLSQRKKEMARRAVNFVVEISRKTGLSNV
 25 IVNIEDLNVRI FHGGGKQAPGWDGFFRPKSENRFIQAIHKAFSDLAHHGIPVIESDPQRT
 SMTCPGCHGCD SKNRNGVRF LCKGCGASMDADFDAACRNLERVALTGKMPKPKSTSCERLLS
 ATTGKVCSDHSLSHDAIEKAS*

>CasΦ-5

MSSLPTPLELLKQKHADLFKGLQFSSKDNKMAGKVLKKGEEAALFLSERGVSRGELPNFR
 30 PPAKTLVVAQSRPFEEFPIYRVSEAIQLYVYSLSVKELETVPSGSSTKKEHQRFQDSSVPD
 FGYTSVQGLNKIFGLARGIYLGVITRGENQLQAKSKHEALNKKRRASGEAETEFDPPTYEY
 MTPERKLAKPPGVNHSIMCYVDISVDEDFRNPDGIVLPSEYAGYCREINTAIEKGTVDRLG
 HLKGGPGYIPGHQRKESTTEGPKINFRKGRIRRSYTALYAKRDSRRVRQGKLALPSYRHMM
 RLNSNAESAILAVIFFGKDWWVFDLRGLLRNVRWRNLFVDGSTPSTLLGMFGDPVIDPKRGV

VAFCYKEQIVPVVSKSITKMVKAPELLNKLKLYLKSEDPLVLVAIDLQTNPNVGVGVYRVMNAS
 LDYEVVTRFALESELLREIESYRQRTNAFEAQIRAETFDAMTSEEQEEITRVRAFSASKAKE
 NVCHRFGMPVDAVDWATMGSENTIHIKWMVRHGDPSLVEVLEYRKDNEIKLDKNGVPKKVKL
 TDKRIANLTSIRLRFSSQETSKHYNDTMWELRRKHPVYQKLSKSKADFSRRVNSIIRRVNHL
 5 VPRARIVFIIEDLKNLKGVFHSGSGKRELGWDSYFEPKSENRFIQLVHKAFSETGKHKGYYI
 IECWPNWTSTCTCPKSCCDSENRHGEVFRCLACGYTCNTDFGTAPDNLVKIATTGKGLPGPK
 KRCKGSSKGKNPKIARSSETGVSVTESGAPKVKKSSPTQTSQSSSQSAP*

>CasΦ-6

MNKIEKEKTPLAKLMNENFAGLRFPFAIIKQAGKKLLKEGELKTIEYMTGKGSIEPLPNFKP
 10 PVKCLIVAKRRDLKYFPICKASCEIQSYVYSLNYKDFMDYFSTPMTSQKQHEEFFKKSGLNI
 EYQNVAGLNLI FNNVKNTYNGVILKVKNRNEKLKKKAIKNNYEFEEIKTFNDDGCLINKPGI
 NNVIYCFQSSISPKILKNITHLPKEYNDYDCSVDRNIIQKYVSRLDIPESQPGHVP EWQRKLP
 EFNNTNNPRRRRKWYSNGRNISKGYSDQVNQAKIEDSLLAQIKIGEDWIIIDIRGLLRDLN
 RRELISYKNKLTIKDVLGFFSDYPIIDIKKNLVTFCYKEGVIQVVSQKSIGNKKSKQLLEKL
 15 IENKPIALV SIDLGQTNPVSVKISKLNKINNKISIESFTYRFLNEEILKEIEKYRKDYDKLE
 LKLINEA

>CasΦ-7

MSNTAVSTREHMSNKTTPPSPLSLLLRAHFPGLKFESQDYKIAGKKLRDGGPEAVISYLTGK
 GQAKLKDVKPPAKAFVIAQSRPFIEWDLVRVSRQIQEKIFGIPATKGRPKQDGLSETAFNEA
 20 VASLEVDGKSKLNEETRAAFYEVGLDAPSLHAQAQNALIKSAISIREGLVKKVENRNEKNL
 SKTKRRKEAGEEATFVEEKAHDERGYLIHPPGVNQTIPGYQAVVIKSCPSDFIGLPSGCLAK
 ESAEALTDYLP HDRMTIPKGQPGYVPEWQHPLLNRKNRRRRDWYSASLNKPKATCSKRSGT
 PNRKNSRTDQIQSGRFKGAIPVLMRFQDEWVIIDIRGLLRNARYRKLLKEKSTIPDLLSLFT
 GDPSIDMRQGVCTFIYKAGQACSAKMVKTNAPEILSELTKSGPVVLV SIDLGQTNPIAAKV
 25 SRVTQLSDGQLSHETLLRELLSNDSSDGKEIARYRVASDRLRDKLANLAVERLSPEHKSEIL
 RAKNDTPALCKARVCAALGLNPEMIAWDKMTPYTEFLATAYLEKGGDRKVATLKPKNRPEML
 RRDIFKFGTEGVRIEVSPEAAEAYREAQWDLQRTSPEYLRLSTWKQELTKRI LNQLRHKAAK
 SSQCEVVVMAFEDLNKMMHGNGKWADGGWDAFFIKKRENRFWMQAFHKS LTELGAHKGVPT
 IEVTPHRTSITCTKCGHCDKANRDGERFACQKCGFVAHADLEIATDNIERVALTGKMPKPE
 30 SERSGDAKKS VGARKAAFKPEEDAEEAE*

>CasΦ-8

MIKPTVSQFLTPGFKLIRNHSRTAGLKLKNEGEEACKKFVRENEIPKDECPNFQGGPAIANI
 IAKSREFTEWEIYQSSLAIQEVI FTLPKDKLPEPILKEEWRAQWLSEHGLDTPYKEAAGLN
 LIIKNAVNTYKGVQVKVDNKNKNNLAKINRKNEIAKLNGEQEISFEEIKAFDDKGYLLQKPS

PNKSIYCYQSVSPKPFITSKYHNVNLPEEYIGYRKSNEPIVSPYQFDRLRPIGEPGYVPK
 WQYTFLSKKENKRRKLSKRIKNVSPILGIICIKKDWCVFDMRGLLRNTNHWKYHKPTDSIND
 LFDYFTGDPVIDTKANVVRFRYKMENGIVNYKPVREKKKGKELLENICDQNGSCKLATVDVGQ
 NNPVAIGLFELKKVNGELTKTLISRHPTPIDFCNKITAYRERYDKLESSIKLDAIKQLTSEQ
 5 KIEVDNYNNNFTPQNTKQIVCSKLNINPNDLPWDKMISGTHFISEKAQVSNKSEIYFTSTDK
 GKTKDVMKSDYKWFQDYKPKLSKEVRDALSDIEWRLRRESLEFNKLSKSREQDARQLANWIS
 SMCDVIGIENLVKKNNFFGGSGKREPGWDNFYKPKKENRWWINAIHKALTELSQNGKGRVIL
 LPAMRTSITCPKCKYCDSKNRNGEKFNCCLKGIELNADIDVATENLATVAITAQSMPTCE
 RSGDAKKPVRARKAKAPEFHDKLAPSYTVVLREAV*

10 >CasΦ-9

MRSSREIGDKILMRQPAEKTAQVFRQEVIGTQKLSGGDAKTAGRLYKQGKMEAAREWLLKG
 ARDDVPPNFQPPAKCLVAVSHPFEEWDISKTNHDVQAYIYAQPLQAEGHLNGLSEKWEDTS
 ADQHKLWFEKTGVPDRGLPVQAINKIAKAAVNRAFGVVRKVENRNEKRSRDNRIAEHNREN
 GLTEVVREAPEVATNADGFLHPPGIDPSILSYASVSPVPYNSSKHSFVRLPEEYQAYNVEP
 15 DAPIPQFVVEDRFAIPPGQPGYVPEWQRLKCSSTNKHRRMRQWSNQDYKPKAGRRAKPLEFQA
 HLTRERAKGALLVVMRIKEDWVVFDRGLLRNVEWRKVLSEEAREKLTLLKGLLDLFTGDPVI
 DTKRGIVTFLYKAEITKILSKRTVTKNARDLLLRLTEPGEDGLRREVGLVAVDLGQTHPIA
 AAIYRIGRTSAGALESTVLHRQGLREDQKEKLKEYRKRHTALDSRLRKEAFETLSVEQQKEI
 VTVSGSGAQITKDKVCNYLGVDPSTLPWEKMGSYTHFISDDFLRRGGDPNIVHFDRQPKKGK
 20 VSKKSQRIKRSDSQWVGRMRPRLSQETAKARMEADWAAQNENEYKRLARSKQELARWCVNT
 LLQNTRCITQCDEIVVIEDLNVKSLHGKGAREPGWDNFFTPKTENRWFIIQLHKTFSLEPK
 HRGEHVIEGCPLRTSITCPACSYCDKNSRNGEKFVCVACGATFHADFEVATYNLVRLATTM
 PMPKSLERQGGGEKAGGARKARKKAKQVEKIVVQANANVTMNGASLHSP*

>CasΦ-10

MDMLDTETNYATETPAQQQDYSPKPPKKAQRAPKGFSSKKARPEKKPPKPITLFTQKHFSGVR
 FLKRVIDRASKILKLSESRITITFLEQAIERDGSAPPDVTTPVHNTIMAVTRPFEEWPEVILS
 KALQKHCHYALTKKIKIKTWPKKGPGKKCLAAWSARTKIPLIPGQVQATNGLFDRIGSIYDGV
 EKKVTNRNANKKLEYDEAIKEGRNPAVPEYETAYNIDGTLINKPGYNPNLYITQSRTPRIT
 EADRPLVEKILWQMVEKKTQSRNQARRARLEKAAHLQGLPVPKFVPEKVDRSQKIEIRIIDP
 30 LDKIEPYMPQDRMAIKASQDGHVPYWQRPFLSKRRNRVRAGWGKQVSSIQAWLTGALLVIV
 RLGNEAFLADIRGALRNAQWRKLLKPDATYQSLFNLFNGDPVVNTRTNHLMAYREGVVNIV
 KSRSEFKGRQTRHLLTLLGQGKTVAGVSFDLGQKHAAGLLAAHFGLGEDGNPVFTPIQACFL
 PQRYLDSLNTYRNRYDALTLDMRRQSLALTPAQQQEFADAQRDPGGQAKRACCLKLNLPD
 EIRWDLVSGISTMISDLYIERGGDPRDVHQVETKPKGKRKSEIRILKIRDGKWAYDFRPKI

ADETRKAQREQLWKLQKASSEFERLSRYKINIARAIAIANWALQWGRELSGCDIVI PVLEDLNV
 GSKFFDGGKGKWLGLGWDNRFTPKKENRWFIVKLHKAVAEELAPHRGVPVYEVMPHRTSMTCPAC
 HYCHPTNREGDRFECQSCHVVKNTRDRVAPYNI LRVAVEGKTLDRWQAEKKPQAEPRMIL
 IDNQES*

- 5 The asterisk (*) in the sequences above denotes a STOP codon. Alternatively, CasΦ-1 is also termed Cas12j ortholog 1. Thus, CasΦ-1- CasΦ-10 may also be referred to as Cas12j orthologs 1-10, respectively.

Guide Polynucleotides

- In an embodiment, the guide polynucleotide is a guide RNA. As used herein, the term
 10 “guide RNA (gRNA)” and its grammatical equivalents can refer to an RNA which can be specific for a target DNA and can form a complex with Cas protein. An RNA/Cas complex can assist in “guiding” 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
 15 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
 20 (the PAM or protospacer adjacent motif) to help distinguish self-versus-non-self. Cas9 nuclease sequences and structures are well known to those of skill in the art (see e.g., “Complete genome sequence of an M1 strain of *Streptococcus pyogenes*.” Ferretti, J.J. *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-
 25 607(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). Cas9 orthologs have been described in various species, including, but not limited to, *S. pyogenes* and *S. thermophilus*. Additional suitable Cas9 nucleases and sequences can be apparent to those of skill in the art based on this disclosure,
 30 and such Cas9 nucleases and sequences include Cas9 sequences from the organisms and loci disclosed in Chylinski, Rhun, and Charpentier, “The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems” (2013) *RNA Biology* 10:5, 726-737; the entire contents of

which are incorporated herein by reference. In some embodiments, a Cas9 nuclease has an inactive (*e.g.*, an inactivated) DNA cleavage domain, that is, the Cas9 is a nickase.

In some embodiments, the guide polynucleotide is at least one single guide RNA (“sgRNA” or “gRNA”). In some embodiments, the guide polynucleotide is at least one
 5 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.

The polynucleotide programmable nucleotide binding domain (*e.g.*, a CRISPR-derived domain) of the base editors disclosed herein can recognize a target polynucleotide
 10 sequence by associating with a guide polynucleotide. A guide polynucleotide (*e.g.*, gRNA) is typically single-stranded and can be programmed to site-specifically bind (*i.e.*, via complementary base pairing) to a target sequence of a polynucleotide, thereby directing a base editor that is in conjunction with the guide nucleic acid to the target sequence. A guide polynucleotide can be DNA. A guide polynucleotide can be RNA. As will be appreciated by
 15 one having skill in the art, in a guide polynucleotide sequence uracil (U) replaces thymine (T) in the sequence. In some cases, the guide polynucleotide comprises natural nucleotides (*e.g.*, adenosine). In some cases, the guide polynucleotide comprises 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,
 20 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, a guide polynucleotide may be truncated by 1, 2, 3, 4, etc. nucleotides, particularly at the 5' end. By way of nonlimiting example, a guide polynucleotide of 20 nucleotides in length may be truncated by 1, 2, 3, 4, etc.
 25 nucleotides, particularly at the 5' end.

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). For example, a guide polynucleotide can
 30 comprise a CRISPR RNA (crRNA) and a trans-activating CRISPR RNA (tracrRNA). For example, a guide polynucleotide can comprise one or more trans-activating CRISPR RNA (tracrRNA).

In type II CRISPR systems, targeting of a nucleic acid by a CRISPR protein (*e.g.*, Cas9) typically requires complementary base pairing between a first RNA molecule (crRNA)

comprising a sequence that recognizes the target sequence and a second RNA molecule (trRNA) comprising repeat sequences which forms a scaffold region that stabilizes the guide RNA-CRISPR protein complex. Such dual guide RNA systems can be employed as a guide polynucleotide to direct the base editors disclosed herein to a target polynucleotide sequence.

5 In some embodiments, the base editor provided herein utilizes a single guide polynucleotide (*e.g.*, sgRNA). In some embodiments, the base editor provided herein utilizes a dual guide polynucleotide (*e.g.*, dual gRNAs). 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
10 herein. For example, a single guide polynucleotide can be utilized for a cytidine base editor and an adenosine base editor.

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
15 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
20 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
25 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
30 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.

A guide RNA or a guide polynucleotide can comprise two or more RNAs, *e.g.*, CRISPR RNA (crRNA) and transactivating crRNA (tracrRNA). A guide RNA or a guide polynucleotide can sometimes comprise a single-chain RNA, or single guide RNA (sgRNA) formed by fusion of a portion (*e.g.*, a functional portion) of crRNA and tracrRNA. A guide RNA or a guide polynucleotide can also be a dual RNA comprising a crRNA and a tracrRNA. Furthermore, a crRNA can hybridize with a target DNA.

As discussed above, a guide RNA or a guide polynucleotide can be an expression product. For example, a DNA that encodes a guide RNA can be a vector comprising a sequence coding for the guide RNA. A guide RNA or a guide polynucleotide can be transferred into a cell by transfecting the cell with an isolated guide RNA or plasmid DNA comprising a sequence coding for the guide RNA and a promoter. A guide RNA or a guide polynucleotide can also be transferred into a cell in other way, such as using virus-mediated gene delivery.

A guide RNA or a guide polynucleotide can be isolated. For example, a guide RNA can be transfected in the form of an isolated RNA into a cell or organism. A guide RNA can be prepared by *in vitro* transcription using any *in vitro* transcription system known in the art. A guide RNA can be transferred to a cell in the form of isolated RNA rather than in the form of plasmid comprising encoding sequence for a guide RNA.

A guide RNA 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 guide RNA can also be different such that each guide RNA guides a fusion protein to a specific target site. Further, second and third regions of each guide RNA can be identical in all guide RNAs.

A first region of a guide RNA or a guide polynucleotide can be complementary to sequence at a target site in a chromosomal sequence such that the first region of the guide

RNA can base pair with the target site. In some cases, a first region of a guide RNA 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,
 5 a region of base pairing between a first region of a guide RNA 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. In some embodiments, a first region of a guide RNA can be or can be about 19, 20, or 21 nucleotides in length.

A guide RNA or a guide polynucleotide can also comprise a second region that forms
 10 a secondary structure. For example, a secondary structure formed by a guide RNA 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
 15 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 guide RNA 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
 20 complementarity to the rest of a guide RNA. 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 guide RNA 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
 25 3 or 4 of a gene. A composition can comprise multiple guide RNAs that all target the same exon or in some cases, multiple guide RNAs that can target different exons. An exon and an intron of a gene can be targeted.

A guide RNA or a guide polynucleotide can target a nucleic acid sequence of or of about 20 nucleotides. A target nucleic acid can be less than or less than about 20 nucleotides.
 30 A target nucleic acid can be at least or at least about 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, or anywhere between 1-100 nucleotides in length. A target nucleic acid can be at most or at most about 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 40, 50, or anywhere between 1-100 nucleotides in length. A target nucleic acid sequence can be or can be about

20 bases immediately 5' of the first nucleotide of the PAM. A guide RNA 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.

A guide polynucleotide, for example, a guide RNA, can refer to a nucleic acid that
 5 can hybridize to another nucleic acid, for example, the target nucleic acid or protospacer in a genome of a cell. A guide polynucleotide can be RNA. A guide polynucleotide can be DNA. The guide polynucleotide can be programmed or designed to bind to a sequence of nucleic acid site-specifically. A guide polynucleotide can comprise a polynucleotide chain and can be called a single guide polynucleotide. A guide polynucleotide can comprise two
 10 polynucleotide chains and can be called a double guide polynucleotide. A guide RNA can be introduced into a cell or embryo as an RNA molecule. For example, a RNA molecule can be transcribed *in vitro* and/or can be chemically synthesized. An RNA can be transcribed from a synthetic DNA molecule, *e.g.*, a gBlocks® gene fragment. A guide RNA can then be introduced into a cell or embryo as an RNA molecule. A guide RNA can also be introduced
 15 into a cell or embryo in the form of a non-RNA nucleic acid molecule, *e.g.*, DNA molecule. For example, a DNA encoding a guide RNA can be operably linked to promoter control sequence for expression of the guide RNA 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 guide RNA include, but
 20 are not limited to, px330 vectors and px333 vectors. In some cases, a plasmid vector (*e.g.*, px333 vector) can comprise at least two guide RNA-encoding DNA sequences.

Methods for selecting, designing, and validating guide polynucleotides, *e.g.*, guide RNAs 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
 25 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 ssDNA 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
 30 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.

5 As a non-limiting example, target DNA hybridizing sequences in crRNAs of a guide RNA for use with Cas9s may be identified using a DNA sequence searching algorithm. gRNA design may be 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
10 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
15 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
20 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 guide RNAs, e.g., crRNAs, may be 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
25 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
30 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.

In some embodiments, a reporter system may be used for detecting base-editing activity and testing candidate guide polynucleotides. In some embodiments, a reporter system may comprise 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.

The guide polynucleotides can be synthesized chemically, synthesized enzymatically, or a combination thereof. For example, the guide RNA can be synthesized using standard phosphoramidite-based solid-phase synthesis methods. Alternatively, the guide RNA can be synthesized in vitro by operably linking DNA encoding the guide RNA 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 guide RNA comprises two separate molecules (e.g., crRNA and tracrRNA), the crRNA can be chemically synthesized and the tracrRNA can be enzymatically synthesized.

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 gRNA, 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.

A DNA sequence encoding a guide RNA or a guide polynucleotide can also be part of a vector. 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 guide RNA can also be linear. A DNA molecule encoding a guide RNA or a guide polynucleotide can also be circular.

In some embodiments, one or more components of a base editor system may be encoded by DNA sequences. 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 guide RNA 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 guide RNA 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 guide RNA).

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, 5'-5' modifications, abasic, acridine, azobenzene, biotin, biotin BB, biotin TEG, cholesteryl TEG, desthiobiotin TEG, DNP TEG, DNP-X, DOTA, dT-Biotin, 5 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 10 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 15 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 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 20 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- 25 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.

30 ***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.

5 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
 10 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),
 15 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
 20 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),
 25 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 1E**

Table 1E. Cas9 proteins and corresponding PAM sequences

Variant	PAM
spCas9	NGG
spCas9-VRQR	NGA
spCas9-VRER	NGCG
SpCas9-MQKFRAER	NGC
xCas9 (sp)	NGN
saCas9	NNGRRT
saCas9-KKH	NNNRRT
spCas9-MQKSER	NGCG
spCas9-MQKSER	NGCN
spCas9-LRKIQK	NGTN
spCas9-LRVSQK	NGTN
spCas9-LRVSQL	NGTN
SpyMacCas9	NAA
Cpf1	5' (TTTV)

In some embodiments, the PAM is NGC. In some embodiments, the NGC PAM is recognized by a Cas9 variant, e.g., an SpCas9 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 2 and 3 below.

Table 2: NGT PAM Variant Mutations at residues 1219, 1335, 1337, 1218

Variant	E1219V	R1335Q	T1337	G1218
1	F	V	T	
2	F	V	R	

Variant	E1219V	R1335Q	T1337	G1218
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 3: 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	

Variant	D1135L	S1136R	G1218S	E1219V	R1335Q
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 Tables 2 and 3. In some embodiments, the variants have improved NGT PAM recognition.

- 5 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 4 below.

Table 4: 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

- 10 In some embodiments, the NGT PAM is selected from the variants provided in Table 5 below.

Table 5. 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.

5 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
10 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,
15 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
20 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
25 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
30 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 embodiments, the Cas9 domains of any of the fusion proteins provided herein
5 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 a Cas9 polypeptide described herein. In some embodiments, the Cas9 domains of any of the fusion proteins provided herein comprises the amino acid sequence of any Cas9 polypeptide described herein. In some
10 embodiments, the Cas9 domains of any of the fusion proteins provided herein consists of the amino acid sequence of any Cas9 polypeptide described 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
15 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,
20 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
25 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
30 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
 5 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
 10 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:

The amino acid sequence of an exemplary PAM-binding SpCas9 is as follows:

MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
 15 RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFFHRLEESFLVEEDKKHERHPIFGNIVD
 EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIT
 QLVQTYNQLFEEINPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGL
 TPNFKNFDFLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLEPKYKEIFFDQSKNGYAGYIDGGASQEEF
 20 YKFIKPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLK
 DNREKIEKILTFRIPIYVGPLARGNSRFAMWTRKSEETITPWNFEVVVDKGASAQSFIERMT
 NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKI IKDKDFLDNEENEDILEDIV
 LTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
 25 LKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVV
 DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEIEGIELGSQILKEHPVENTQL
 QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
 NVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
 VAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
 30 VGTALIKKYPKLESEFVYGDKYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
 GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKRNS
 DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEKPNP
 IDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNELALPSKYVNFYLYLAS
 HYEKLKGSPEQKQLFVEQHKHYLDEIIIEQISEFSKRVILADANLDKVL SAYNKHRRDKPI

REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDL SQ
LGGD.

The amino acid sequence of an exemplary PAM-binding SpCas9n is as follows:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
5 RLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFI
QLVQTYNQLFEE NPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLI ALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP EKYKEIFFDQSKNGYAGYIDGGASQEEF
10 YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLK
DNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
15 LKSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA IKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKMKMNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
20 VGTALIKKYPKLESEFVYG DYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPK KYGGFDSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSSFENP
IDFLEAKGYKEVKKDLIIKLPKYSLFEL ENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGS PEDNEQKQLFVEQH KHYLDEII EQISEFSKRVILADANLDKVL SAYNKH RDKPI
25 REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDL SQ
LGGD.

The amino acid sequence of an exemplary PAM-binding SpEQR Cas9 is as follows:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHRLEESFVEEDKKHERHPIFGNIVDE
30 VAYHEKYPTIYHLRKKLVDSTDKADLRLLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQ
LVQTYNQLFEE NPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLI ALSLGLT
PNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTE
ITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP EKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKD

NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASAQSFIERMTN
 FDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV
 TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVL
 TLTFLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFL
 5 KSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAICKGILQTVKVVD
 ELVKVMGRHKPENIVIMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
 NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDN
 VPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV
 AQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
 10 GTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANG
 EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSD
 KLIARKKDWDPKKYGGF**E**SPTVAYSVLVVAKEKGKSKKLKSVKELLGITIMERSSEFKNPI
 DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
 YEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKVLSAYNKHDKPIR
 15 EQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**QYR**STKEVLDTLIHQSIITGLYETRIDLSQL
 GGD. In this sequence, residues E1135, Q1335 and R1337, which can be mutated from
 D1135, R1335, and T1337 to yield a SpEQR Cas9, are underlined and in bold.

The amino acid sequence of an exemplary PAM-binding SpVQR Cas9 is as follows:
 MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
 20 RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVD
 EVAYHEKYPTIYHLRKKLVDSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFI
 QLVQTYNQLFEEFNINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGL
 TPNFKNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPKEYKEIFFDQSKNGYAGYIDGGASQEEF
 25 YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIHQIHLGELHAILRRQEDFYFPLK
 DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASAQSFIERMT
 NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
 LTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
 30 LKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAICKGILQTVKV
 DELVKVMGRHKPENIVIMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
 QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
 NVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
 VAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAV

VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
 GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNS
 DKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKEKGKSKKLKSVKELLGITIMERSSEKPN
 IDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
 5 HYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFSKRVILADANLDKVL SAYNKH RDKPI
 REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL DATLIHQSI TGLYETRIDL SQ
 LGGD. In this sequence, residues V1135, Q1335, and R1337, which can be mutated from
 D1135, R1335, and T1337 to yield a SpVQR Cas9, are underlined and in bold.

The amino acid sequence of an exemplary PAM-binding SpVRER Cas9 is as follows:

10 MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
 RLKRTARRRYTRKRNRI CYLQEIFS NEMAKVDDSFHRL EESFLVEEDKKHERHPIFGNIVD
 EVAYHEKYPTIYHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFI
 QLVQTYNQLF EENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGL
 TPNFKNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
 15 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP EKYKEIFFDQSKNGYAGYIDGGASQEEF
 YKFIKPILEKMDGTEELLV KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLK
 DNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWNFEVV DKGASAQSFIERMT
 NFDKNLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIV
 20 LTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
 LKSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA IKKGILQTVKVV
 DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
 QNEKLYLYYLQNGRDMYVDQELDINRLSDYDV DHIVPQSFLKDDSIDKNVLTRSDKNRGKSD
 NVPSEEVVKMKMNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
 25 VAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
 VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
 GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNS
 DKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKEKGKSKKLKSVKELLGITIMERSSEKPN
 IDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASARELQKGNELALPSKYVNFLYLAS
 30 HYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFSKRVILADANLDKVL SAYNKH RDKPI
 REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKEYRSTKEVL DATLIHQSI TGLYETRIDL SQ
 LGGD. In the above sequence, residues V1135, R1218, Q1335, and R1337, which can be
 mutated from D1134, G1218, R1335, and T1337 to yield a SpVRER Cas9, are underlined
 and in bold.

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 1A-1E**). 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 Jakimo *et al.*, (www.biorxiv.org/content/biorxiv/early/2018/09/27/429654.full.pdf), and is provided below.

SpyMacCas9

```
MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAET
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFI
QLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLLENLIAQLPGEKRNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNS
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLEPKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIHQIHLGELHAILRRQEDFYFPLK
DNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASAQSFIERMT
NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDRGMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGHSLEHQIANLAGSPAIKKGILQTVKIV
DELVKVMGHKPENIVIMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIDKDDSIDNKVLTRSDKNRGKSDN
```

VPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV
 AQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
 GTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANG
 EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIIVKKTEIQTGQNGGLFDDNPKSP
 5 LEVTPSKLVPLKKEINPKKYGGYQKPTTAYPVLLITDTKQLIPISVMNKKQFEQNPVKFLRD
 RGYQQVGKNDFIKLPKYTLVDIGDGIKRLWASSKEIHKGNQLVVSCKSQILLYHAHHLDS
 SLDYLNQHNQQFDVLFNEIISFSKKCKLGKEHIQKIENVYSNKKNSASIEELAESFIKLLGF
 TQLGATSPFNFLGVKLNQKQYKGGKDYILPCTEGTLIRQSITGLYETRVDLSKIGED.

In some cases, a variant Cas9 protein harbors, H840A, P475A, W476A, N477A,
 10 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 cases, the variant
 Cas9 protein harbors D10A, H840A, P475A, W476A, N477A, D1125A, W1126A, and
 15 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 cases, when a variant Cas9 protein harbors W476A and W1126A mutations or when
 the variant Cas9 protein harbors P475A, W476A, N477A, D1125A, W1126A, and D1218A
 20 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 cases, 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
 25 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.

30 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; S. 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. By way of example, S. Miller *et al.* (2020, *Id.*) describes SpCas9 variants that collectively recognize Non-G PAMs, such as NRNH PAMs (where R is A or G and H is A, C or T).

Fusion proteins comprising a Cas9 domain 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 and one or more adenosine deaminase domain, cytidine deaminase domain, and/or DNA glycosylase 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 adenosine deaminases and cytidine deaminases described 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:

NH₂-[A-B-C]-COOH;
NH₂-[A-B-C-D]-COOH; or
NH₂-[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,
a DNA glycosylase 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:

NH₂-[A_n-B_o-C_n]-COOH;

NH₂-[A_n-B_o-C_n-D_o]-COOH; or

NH₂-[A_n-B_o-C_p-D_o-E_q]-COOH;

5 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,

a DNA glycosylase 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

10 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;

15 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;

20 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
25 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;

30 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.

5 In some embodiments, the adenosine deaminase of the fusion protein comprises a
 TadA*8 and a cytidine 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
 10 TadA*8.24. In some embodiments, the adenosine deaminase of the fusion protein comprises
 a TadA*9 and a cytidine deaminase.

Exemplary fusion protein structures include the following:

NH₂-[TadA*8]-[Cas9 domain]-COOH;
 NH₂-[Cas9 domain]-[TadA*8]-COOH;
 15 NH₂-[TadA*8]-[Cas12 domain]-COOH;
 NH₂-[Cas12 domain]-[TadA*8]-COOH;
 NH₂-[TadA*9]-[Cas9 domain]-COOH;
 NH₂-[Cas9 domain]-[TadA*9]-COOH;
 NH₂-[TadA*9]-[Cas12 domain]-COOH;
 20 NH₂-[Cas12 domain]-[TadA*9]-COOH;
 NH₂-[adenosine deaminase]-[Cas9/12]-[cytidine deaminase]-COOH;
 NH₂-[cytidine deaminase]-[Cas9/12]-[adenosine deaminase]-COOH;
 NH₂-[TadA*8]-[Cas9/12]-[cytidine deaminase]-COOH;
 NH₂-[cytidine deaminase]-[Cas9/12]-[TadA*8]-COOH;
 25 NH₂-[TadA*9]-[Cas9/12]-[cytidine deaminase]-COOH; or
 NH₂-[cytidine deaminase]-[Cas9/12]-[TadA*9]-COOH.

In some embodiments, the fusion proteins comprising a cytidine deaminase, abasic
 editor, and/or adenosine deaminase and a napDNAbp (e.g., Cas9 domain) do not include a
 linker sequence. In some embodiments, a linker is present between the cytidine deaminase
 30 and adenosine deaminase domains and the napDNAbp. In some embodiments, the “-” used in
 the general architecture above 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.

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 in 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, any of the fusion proteins provided herein further comprise a nuclear localization sequence (NLS). In some embodiments, the NLS is fused to the N-terminus of the fusion protein. In some embodiments, the NLS is fused to the C-terminus of the fusion protein. In some embodiments, the NLS is fused to the N-terminus of the Cas9 domain. In some embodiments, the NLS is fused to the C-terminus of an nCas9 domain or a dCas9 domain. In some embodiments, the NLS is fused to the N-terminus of the deaminase. In some embodiments, the NLS is fused to the C-terminus of the 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, KRTADGSEFESPKKKRKV, KRPAATKKAGQAKKKK, KKTELQTTNAENKTKKL, KRGINDRNFWRGENGRKTR, RKSGKIAAIVVKRPRKPKKKRKV, or MDSELLMNRKFLYQFKNVRWAKGRRETYLC.

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), NH₂ is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein:

NH₂-NLS-[cytidine deaminase]-[napDNAbp domain]-COOH;
 NH₂-NLS [napDNAbp domain]-[cytidine deaminase]-COOH;
 NH₂-[cytidine deaminase]-[napDNAbp domain]-NLS-COOH;
 NH₂-[napDNAbp domain]-[cytidine deaminase]-NLS-COOH;
 NH₂-NLS-[adenosine deaminase]-[napDNAbp domain]-COOH;
 NH₂-NLS [napDNAbp domain]-[adenosine deaminase]-COOH;
 NH₂-[adenosine deaminase]-[napDNAbp domain]-NLS-COOH;
 NH₂-[napDNAbp domain]-[adenosine deaminase]-NLS-COOH;
 NH₂-NLS-[cytidine deaminase]-[napDNAbp domain]-[adenosine deaminase]-COOH;

NH₂-NLS-[adenosine deaminase]-[napDNAbp domain]-[cytidine deaminase]-COOH;
 NH₂-NLS-[adenosine deaminase] [cytidine deaminase]-[napDNAbp domain]-COOH;
 NH₂-NLS-[cytidine deaminase]-[adenosine deaminase]-[napDNAbp domain]-COOH;
 NH₂-NLS-[napDNAbp domain]-[adenosine deaminase]-[cytidine deaminase]-COOH;
 5 NH₂-NLS-[napDNAbp domain]-[cytidine deaminase]-[adenosine deaminase]-COOH;
 NH₂-[cytidine deaminase]-[napDNAbp domain]-[adenosine deaminase]-NLS-COOH;
 NH₂-[adenosine deaminase]-[napDNAbp domain]-[cytidine deaminase]-NLS-COOH;
 NH₂-[adenosine deaminase] [cytidine deaminase]-[napDNAbp domain]-NLS-COOH;
 NH₂-[cytidine deaminase]-[adenosine deaminase]-[napDNAbp domain]-NLS-COOH;
 10 NH₂-[napDNAbp domain]-[adenosine deaminase]-[cytidine deaminase]-NLS-COOH; or
 NH₂-[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, the linkers described herein. In some embodiments, the N-terminus or C-terminus NLS is a bipartite NLS. A bipartite NLS comprises two basic amino acid clusters, which are separated by a relatively short spacer sequence (hence bipartite - 2 parts, while
 15 monopartite NLSs are not). The NLS of nucleoplasmin, KR [PAATKKAGQA] KKKK, 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.

20 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 ammo-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 of these (*e.g.*, one or more NLS at the ammo-terminus and one or more NLS
 25 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
 30 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.

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. 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 inserted at an internal location of the napDNAbp.

In some embodiments, the heterologous polypeptide is a 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. The deaminase in a fusion protein can be an adenosine deaminase. In some embodiments, the adenosine deaminase is a TadA (e.g., TadA*7.10, TadA*8 or TadA*9). In some embodiments, the TadA is a TadA*8. TadA sequences (e.g., TadA7.10, TadA*8 or TadA*9) as described herein are suitable deaminases for the above-described fusion proteins.

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 permuted at amino acid residue 116 as numbered in the TadA reference sequence. In some embodiments, the deaminase is a circular permutant TadA, circularly permuted at amino acid residue 136 as numbered in the TadA reference sequence. In some embodiments, the deaminase is a circular permutant TadA, circularly permuted at amino acid residue 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 deaminase. In some embodiments, the fusion protein comprises two deaminases. The two or more deaminases in a fusion protein can be an adenosine deaminase, cytidine deaminase, or a combination thereof. The two or more deaminases can be homodimers. The two or more deaminases can be 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 thereof.

The Cas9 polypeptide of a fusion protein can comprise 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 Cas9 polypeptide.

The Cas9 polypeptide of a fusion protein can comprise 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 the Cas9 amino acid sequence set forth below (called the “Cas9 reference sequence” below):

MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
 RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFFHRLEESFLVEEDKKHERHPIFGNIVD
 EVAYHEKYPTIYHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSD VDKLFI
 QLVQTYNQLF EENPINASGVDAKAILSARLSKSRRL ENLIAQLPGEKKNGLFGNLIALSLGL
 TPNF KSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP EKYKEIFFDQSKNGYAGYIDGGASQEEF
 YKFIKPILEKMDGTEELLV KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLK
 DNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEVV DKGASAQSFIERMT
 NFDKNLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
 LTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
 LKSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA IKKGILQTVKVV
DELVKVMGRHKPENIVIMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYKVREINNYHHAHDAYLNAV

VGTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELNGRKRMLASAGELQKGNELALPSKYVNFLYLAS
5 HYEKLKGS PEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLDKVL SAYNKH RDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDL SQ
LGGD (single underline: HNH domain; double underline: RuvC domain).

Fusion proteins comprising a heterologous catalytic domain flanked by N- and C-
terminal fragments of a Cas9 polypeptide are also useful for base editing in the methods as
10 described herein. Fusion proteins comprising Cas9 and one or more deaminase domains, *e.g.*,
adenosine deaminase, or comprising an adenosine deaminase domain flanked by Cas9
sequences are also useful for highly specific and efficient base editing of target sequences. In
an embodiment, a chimeric Cas9 fusion protein contains a heterologous catalytic domain
(*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine
15 deaminase) inserted within a Cas9 polypeptide. 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,
20 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:

25 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
30 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 or TadA*9. In some embodiments, a TadA*8 or TadA*9 is fused within Cas9 and a cytidine deaminase is fused to the C-terminus. In some embodiments, a TadA*8 or TadA*9 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 or TadA*9 is fused to the C-terminus. In some embodiments, a cytidine deaminase is fused within Cas9 and a TadA*8 or TadA*9 fused to the N-terminus. Exemplary structures of a fusion protein with a TadA*8 or TadA*9 and a cytidine deaminase and a Cas9 are provided as follows:

NH₂-[Cas9(TadA*8 or TadA*9)]-[cytidine deaminase]-COOH;

NH₂-[cytidine deaminase]-[Cas9(TadA*8 or TadA*9)]-COOH;

NH₂-[Cas9(cytidine deaminase)]-[TadA*8 or TadA*9]-COOH; or

NH₂-[TadA*8 or TadA*9]-[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 C α 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 C α 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 napDNAAbp 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, an adenosine deaminase (*e.g.*, TadA*9) 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 adenosine deaminase (*e.g.*, TadA*9) 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 adenosine deaminase (*e.g.*, TadA*9) 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 adenosine deaminase (*e.g.*, TadA*9) 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
5 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
10 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
15 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
20 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
25 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
30 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
5 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
10 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
15 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
20 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
25 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
30 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.

5 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 6 below:

Table 6: 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

10 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

15 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,

20 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
 5 embodiments, the HNH domain is deleted and the deaminase is inserted in its place. In some
 10 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
 10 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-
 15 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
 20 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
 25 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 adenosine 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.

30 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
 5 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
 10 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
 15 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,
 20 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
 25 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.

30 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, (GGGGS)_n, (G)_n, (EAAAK)_n, (GGG)_n, SGSETPGTSESATPES. 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 or GSSGSETPGTSESATPESSG. In other embodiments, the linker is a rigid linker. In other embodiments of the above aspects, the linker is encoded by GGAGGCTCTGGAGGAAGC or GGCTCTTCTGGATCTGAAACACCTGGCACAAGCGAGAGCGCCACCCCTGAGAGCTCTGGC.

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;

5 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.*,
10 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 or TadA*9. In some embodiments, a TadA*8 or TadA*9 is fused within Cas12 and a cytidine deaminase is fused to the C-terminus. In some embodiments, a TadA*8 or
15 TadA*9 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 or TadA*9 is fused to the C-terminus. In some embodiments, a cytidine deaminase is fused within Cas12 and a TadA*8 or TadA*9 fused to the N-terminus. Exemplary structures of a fusion protein with a TadA*8 or TadA*9 and a cytidine deaminase and a Cas12 are provided as follows:

N-[Cas12(TadA*8 or TadA*9)]-[cytidine deaminase]-C;

20 N-[cytidine deaminase]-[Cas12(TadA*8 or TadA*9)]-C;

N-[Cas12(cytidine deaminase)]-[TadA*8 or TadA*9]-C; or

N-[TadA*8 or TadA*9]-[Cas12(cytidine deaminase)]-C.

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

25 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
30 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. In other embodiments, the Cas12 polypeptide has at least about 90% amino acid sequence identity to *Bacillus hisashii* Cas12b, *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
 5 *Bacillus hisashii* Cas12b, *Bacillus thermoamylovorans* Cas12b, *Bacillus sp. V3-13* Cas12b, 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.

10 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
 15 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
 20 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,
 25 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
 30 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/Cas Φ . 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
 5 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
 10 nuclear localization signal is MAPKKKRKVGIIHGVPA. In other embodiments of the above aspects, the nuclear localization signal is encoded by the following sequence: ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCC. 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,
 15 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
 20 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 7 below.

Table 7: 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

BhCas12b	Insertion site	Inserted between aa
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.*, ABE8.13) may be inserted into a BhCas12b to produce a fusion protein (*e.g.*, ABE8.13-BhCas12b) that effectively edits a nucleic acid sequence.

In some embodiments, the base editing system described herein comprises an ABE with TadA inserted into a Cas9. Exemplary sequences of ABEs having a TadA inserted into a Cas9 protein are described in International PCT Application No. PCT/US2020/048586, filed August 28, 2020, the contents of which are incorporated by reference herein in their entirety.

Cas9 Domains with Reduced Exclusivity

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 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. 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); Nishimasu, H., *et al.*, “Engineered CRISPR-Cas9 nuclease with expanded targeting space” *Science*. 2018 Sep 21;361(6408):1259-1262, 5 Chatterjee, P., *et al.*, “Minimal PAM specificity of a highly similar SpCas9 ortholog” *Sci Adv*. 2018 Oct 24;4(10):eaau0766. doi: 10.1126/sciadv.aau0766, the entire contents of each are hereby incorporated by reference.

Nucleobase Editing Domain

10 Described herein are base editors comprising a fusion protein that includes a polynucleotide programmable nucleotide binding domain and a nucleobase editing domain (*e.g.*, a deaminase domain). The base editor can be programmed to edit one or more bases in a target polynucleotide sequence by interacting with a guide polynucleotide capable of recognizing the target sequence. Once the target sequence has been recognized, the base 15 editor is anchored on the polynucleotide where editing is to occur and the deaminase domain components of the base editor can then edit a target base.

In some embodiments, the nucleobase editing domain includes a deaminase domain. In some embodiments, base editors include cytidine base editors (*e.g.*, BE4) that convert target C•G base pairs to T•A. In some embodiments, base editors include adenine base 20 editors (*e.g.*, ABE7.10) that convert A•T to G•C. As particularly described herein, the deaminase domain includes an adenosine deaminase. In some embodiments, the terms “adenine deaminase” and “adenosine deaminase” can be used interchangeably. 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 25 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 30 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 to G Editing

In some embodiments, a base editor described herein can comprise a deaminase domain which includes an adenosine deaminase. 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, the nucleobase editors provided herein can be made by fusing together one or more protein domains, thereby generating a fusion protein. In certain embodiments, the fusion proteins provided herein comprise one or more features that improve the base editing activity (*e.g.*, efficiency, selectivity, and specificity) of the fusion proteins. For example, the fusion proteins provided herein can comprise a Cas9 domain that has reduced nuclease activity. In some embodiments, the fusion proteins provided herein can 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 containing a T opposite the targeted A. Mutation of the catalytic residue (*e.g.*, D10 to A10) of Cas9 prevents cleavage of the edited strand containing the targeted A residue. Such Cas9 variants are able to 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 T to C change on the non-edited strand. 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). In another embodiment, an adenosine deaminase incorporated into a base editor comprises all or a portion of adenosine deaminase acting on 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.

The adenosine deaminase can be derived from any suitable organism (*e.g.*, *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 corresponds to any of the mutations described herein (*e.g.*, any of the mutations identified in ecTadA) can be generated accordingly.

Adenosine deaminases

In some embodiments, fusion proteins described herein can comprise a deaminase domain which includes an adenosine deaminase. 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, the adenosine deaminases provided herein are capable of deaminating adenine. In some embodiments, the adenosine deaminases provided herein are capable of deaminating adenine in a deoxyadenosine residue of DNA. 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). 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 adenosine deaminase (*e.g.*, having homology to ecTadA) that corresponds to any of the mutations described herein, *e.g.*, any of the mutations identified in ecTadA. 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*.

The disclosure provides adenosine deaminase variants that have increased efficiency (>50-60%) 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 (*i.e.*, “bystanders”).

In particular embodiments, the TadA is any one of the TadA described in PCT/US2017/045381 (WO 2018/027078), which is incorporated herein by reference in its entirety. The wild-type TadA (TadA(wt)) or “the TadA reference sequence” is as follows:

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHP
GMNHRVEITEGILADECAALLSDFFRMRQEIKAKKQKQSSTD

In some embodiments, the nucleobase editors of the disclosure are adenosine deaminase variants comprising an alteration in the following sequence:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNNRVIGEGWNRRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (also termed TadA*7.10).

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

ABE8 comprising a heterodimer of a TadA*8 variant. In some embodiments, the TadA*8 variant is selected from one or more of Table 8, 10, 11, 12, or 13.

In some embodiments, the base editor is ABE9 comprising a TadA*9 variant. In some embodiments, the base editor is ABE9 comprising a TadA*9 variant monomer. In some embodiments, the base editor is ABE9 comprising a heterodimer of a TadA*9 variant and a TadA(wt). In some embodiments, the base editor is ABE9 comprising a heterodimer of a TadA*9 variant and another TadA variant (*e.g.*, TadA*7.10). In some embodiments, the base editor is ABE9 comprising a homodimer of a TadA*9 variant. In some embodiments, the TadA*9 variant is as provided in Tables 14 and 18 herein. In some embodiments, the TadA*9 variant is selected from the variants described below and with reference to the following sequence (termed TadA*7.10):

```

      10      20      30      40      50
MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG
      60      70      80      90     100
15 LHDPTAHA EI MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG
      110     120     130     140     150
RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCYFFR
      160
MPRQVFNAAQK KAQSSTD.

```

In some embodiments, an adenosine deaminase (*e.g.*, TadA*9) comprises an alteration at an amino acid position selected from the group consisting of 21, 23, 25, 38, 51, 54, 70, 71, 72, 73, 94, 124, 133, 139, 146, and 158 of SEQ ID NO: 1, or a corresponding alteration in another adenosine deaminase. In some embodiments, an adenosine deaminase (*e.g.*, TadA*9) 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, or a corresponding alteration in another adenosine deaminase. The relevant bases altered in the reference sequence are shown by 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; 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; 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 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, 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 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 some embodiments the TadA deaminase is a full-length *E. coli* TadA deaminase. For example, in certain embodiments, the adenosine deaminase comprises the amino acid sequence:

```
MRRAFITGVFFLSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGR
HDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFSGARDAKTGA
AGSLMDVLHHPGMNRHVEITEGILADECAALLSDFFRMRQEIKAQKKAQSSTD.
```

It should be appreciated, however, that additional adenosine deaminases useful in the present application would be apparent to the skilled artisan and are within the scope of this disclosure. For example, the adenosine deaminase may be a homolog of adenosine deaminase acting on tRNA (ADAT). Without limitation, the amino acid sequences of exemplary AD AT homologs include the following:

Staphylococcus aureus TadA:

MGSHMTNDIYFMTLAIIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLLQQPTAHAEHIA
IERAAKVLGSRLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCSGS
LMNLLQQSNFNHRAIVDKGVLKEACSTLLTFFKNLRANKKSTN

5 *Bacillus subtilis* TadA:

MTQDELYMKEAIKEAKKAEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEMLVIDEA
CKALGTWRLEGATLYVTLEPCPMCAGAVVLSRVEKVVFGAFDPKGGCSGTLMNLLQEERFNH
QAEVVSGLVEECCGMLS AFFREL RKKKKAARKNLSE

Salmonella typhimurium (*S. typhimurium*) TadA:

10 MPPAFITGVTSLSDVELDHEYWMRHALTLAKRAWDEREVPVGAVLVHNHRVIGEGWNRPIGR
HDPTAHAEIMALRQGGVLVQNYRLD TTLYVTLEPCVMCAGAMVHSRI GRVVFGARDAKTGA
AGSLIDVLHHPGMNHRVEIIIEGVL RDECATLLSDFFRMRRQEIKALKKADRAEGAGPAV

Shewanella putrefaciens (*S. putrefaciens*) TadA:

MDEYWMQVAMQMAEKAEAAAGEVPVGAVLVKDGQQIATGYNLSISQHDPTAHAEILCLRSAGK
15 KLENYRLLDATLYITLPCAMCAGAMVHSRIARVVYGARDEKTGAAGTVVNLLQHAPFNHQQV
EVTSGVLAEACSAQLSRFFKRRRDEKKALKLAQRAQQGIE

Haemophilus influenzae F3031 (*H. influenzae*) TadA:

MDAAKVRSEFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQSDPTAHA
EIIALRNGAKNIQNYRLNSTLYVTLEPCTMCAGAILHSRIKRLVFGASDYKTGAIGSRFHF
20 FDDYKMNHTLEITSGVLAEESQKLSTFFQKRREEKKIEKALLKSLSDK

Caulobacter crescentus (*C. crescentus*) TadA:

MRTDESEDQDHRMRLALDAARAAAEAGETPVGAVILDPSTGEVIATAGNGPIAAHDPTAHA
EIAAMRAAAAKLGNRYRLD TLTVVTLEPCAMCAGAI SHARIGRVVFGADDPKGGAVVHGPKF
FAQPTCHWRPEVTGGVLADESADLLRGFFRARRKAKI

25 *Geobacter sulfurreducens* (*G. sulfurreducens*) TadA:

MSSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRGNLREGSNDPSAHA
EMIAIRQAARRSANWRLTGATLYVTLEPCLMCMGAIILARLERVVFGCYDPKGAAGSLYDL
SADPRLNHQVRLSPGVCQEECGTMLSDFFRDLRRRKAKATPALFIDERKVPPEP

An embodiment of *E. Coli* TadA (ecTadA) includes the following:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
 GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

In some embodiments, the adenosine deaminase is from a prokaryote. In some
 5 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 one embodiment, a fusion protein of the disclosure comprises a wild-type TadA
 10 linked to TadA*7.10, which is linked to Cas9 nickase. In particular embodiments, the fusion
 proteins comprise a single TadA*7.10 domain (e.g., provided as a monomer). In other
 embodiments, the ABE7.10 editor comprises TadA*7.10 and TadA(wt), which are capable of
 forming heterodimers.

In some embodiments, the adenosine deaminase comprises an amino acid sequence
 15 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
 20 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, 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 mutations compared to a reference
 25 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
 30 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
5 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 (*e.g.*,
10 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 a D108G, D108N, D108V, D108A, or D108Y mutation, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an A106X mutation in
15 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 A106V mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, wild type TadA or ecTadA).

In some embodiments, the adenosine deaminase comprises a E155X mutation in
20 TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA), 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
25 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 (*e.g.*,
ecTadA), 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
30 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.

5 For example, an adenosine deaminase can 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 (*e.g.*,
 10 ecTadA): 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 can be made in an adenosine deaminase (*e.g.*,
 15 ecTadA).

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
 20 corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA), 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
 25 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 (*e.g.*, ecTadA).

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
 30 mutations in another adenosine deaminase (*e.g.*, ecTadA), 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 (*e.g.*, ecTadA).

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 (*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 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 (*e.g.*, ecTadA).

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).

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
 5 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”
 10 Nature, 551, 464-471 (2017), the entire contents of which are hereby incorporated by reference.

In some embodiments, the adenosine deaminase comprises one or more corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA). In some
 15 embodiments, the adenosine deaminase comprises a D108N, D108G, or D108V mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises a A106V and D108N mutation in TadA reference sequence, or corresponding mutations in another adenosine
 20 deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises R107C and D108N mutations in TadA reference sequence, or corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises a H8Y, D108N, N127S, D147Y, and Q154H mutation in TadA reference
 25 sequence, or corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA). 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 (*e.g.*, ecTadA). In some
 30 embodiments, the adenosine deaminase comprises a D108N, D147Y, and E155V mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA). In some
 35 embodiments, the adenosine deaminase comprises a H8Y, D108N, and N127S mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA). In some
 40 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 a 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 (*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 H123Y mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an I156X mutation in 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 I156F mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

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 (*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 S2X, I49X, A106X, D108X, D147X, and E155X 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, A106X, D108X, N127X, and K160X 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, 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 (*e.g.*, ecTadA). 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 (*e.g.*, ecTadA).

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 (*e.g.*, ecTadA), 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 (*e.g.*, ecTadA). 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 (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an E25X mutation in 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 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 (*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 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 (*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 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 (*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 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 (*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 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 (*e.g.*, ecTadA), 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).

5 In some embodiments, the adenosine deaminase comprises an H36X mutation in 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 H36L mutation in TadA reference sequence, or a corresponding mutation in another
10 adenosine deaminase (*e.g.*, ecTadA).

 In some embodiments, the adenosine deaminase comprises an N37X mutation in 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
15 an N37T, or N37S mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

 In some embodiments, the adenosine deaminase comprises an P48X mutation in 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
20 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 (*e.g.*, ecTadA).

 In some embodiments, the adenosine deaminase comprises an R51X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase,
25 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 (*e.g.*, ecTadA).

 In some embodiments, the adenosine deaminase comprises an S146X mutation in
30 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 S146R, or S146C mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an K157X mutation in 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 a K157N mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an P48X mutation in 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 a P48S, P48T, or P48A 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 (*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 a A142N mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an W23X mutation in 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 a W23R, or W23L mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an R152X mutation in 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 a R152P, or R52H mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

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

5 " _ " and each combination of mutations is between parentheses:

(A106V_D108N),

(R107C_D108N),

(H8Y_D108N_N127S_D147Y_Q154H),

(H8Y_D108N_N127S_D147Y_E155V),

10 (D108N_D147Y_E155V),

(H8Y_D108N_N127S),

(H8Y_D108N_N127S_D147Y_Q154H),

(A106V_D108N_D147Y_E155V),

(D108Q_D147Y_E155V),

15 (D108M_D147Y_E155V),

(D108L_D147Y_E155V),

(D108K_D147Y_E155V),

(D108I_D147Y_E155V),

(D108F_D147Y_E155V),

20 (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),

25 (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),

30 (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

35 _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
 5 _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),
 10 (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),
 15 (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),
 20 (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),
 25 (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),
 30 (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),
 35 (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),
 5 (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),
 10 (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),
 15 (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
 20 _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
 25 _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),
 30 (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
 35 _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),
5 (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_E155V
10 _I156F_K157N).

In certain embodiments, the fusion proteins provided herein comprise one or more features that improve the base editing activity of the fusion proteins. For example, any of the fusion proteins provided herein may comprise a Cas9 domain that has reduced nuclease activity. In some embodiments, any of the fusion proteins provided herein may have a Cas9
15 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).

In some embodiments, the adenosine deaminase is TadA*7.10. In some embodiments, TadA*7.10 comprises at least one alteration. In particular embodiments, TadA*7.10 comprises one or more of the following alterations: Y147T, Y147R, Q154S,
20 Y123H, V82S, T166R, and Q154R. The alteration Y123H is also referred to herein as H123H (the alteration H123Y in TadA*7.10 reverted back to Y123H (wt)). In other embodiments, the 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 +
25 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 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
30 sequence, or a corresponding mutation in another TadA.

In other embodiments, a base editor of the disclosure is a monomer comprising an adenosine deaminase variant (*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 (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, a base editor is a heterodimer comprising a wild-type adenosine deaminase and an adenosine deaminase variant (*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 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: 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 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 base editor is a heterodimer comprising a wild-type adenosine deaminase 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 is a heterodimer comprising a TadA*7.10 domain and an adenosine deaminase variant (*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 base editor is a heterodimer comprising 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 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
```

In some embodiments, the TadA*8 is a 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, 16, 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, 16, 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 is a monomer comprising an adenosine deaminase variant (*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 adenosine deaminase variant (TadA*8) is a monomer 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 is a heterodimer comprising a wild-type adenosine deaminase and an adenosine deaminase variant (*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 is a heterodimer comprising a wild-type adenosine deaminase 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 is a heterodimer comprising a TadA*7.10 domain and an adenosine deaminase variant (*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 is a heterodimer comprising 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 some embodiments, the TadA*8 is a variant as shown in Table 8. Table 8 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 8 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 8. Additional TadA*8 Variants

TadA	TadA amino acid number									
	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 one embodiment, a fusion protein of the disclosure comprises a wild-type TadA 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. Exemplary sequences follow:

TadA(wt) or “the TadA reference sequence”:

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHP
GMNHRVEITEGILADECAALLSDFFRMRQEIKAKKQKSSTD

TadA*7.10:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQKSSTD

TadA*8:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
 GMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSD.

In some embodiments, the adenosine deaminase comprises an amino acid sequence
 5 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
 10 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, 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 mutations compared to a reference
 15 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
 20 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
25	LHDPTAHAEI	MALRQGGLVM	QNY <u>Y</u> RLIDATL	<u>Y</u> VTFEPCVMC	AGAMIHSRIG	100
	RVVFGVRNAK	TGAAGSLMDV	LH <u>Y</u> PGMNHRV	EITEGILADE	CAALLC <u>Y</u> FFR	150
	MPR <u>Q</u> VFNAQK	KAQSS <u>T</u> D				

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,
 30 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.

5 In some embodiments, the adenosine deaminase is TadA*8, which comprises or consists essentially of the following sequence or a fragment thereof having adenosine deaminase activity:

MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG
LHDPTAHAEI MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIH SRIG
10 RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCTFFR
MPRQVFNAQK KAQSSTD

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 disclosure 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.

Cas9 complexes with guide RNAs

25 Some aspects of this disclosure provide complexes comprising any of the fusion proteins provided herein, and a guide RNA bound to a Cas9 domain (*e.g.*, a dCas9, a nuclease active Cas9, or a Cas9 nickase) of fusion protein. 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 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 1 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, NGC, NGCG, NGN, NNGRRT, NNNRRT, NGCG, NGCN, NGTN, NGTN, or 5' (TTTV) sequence.

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 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 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
 5 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.

Additional Domains

A base editor described herein can include any domain which helps to facilitate the
 10 nucleobase editing, modification or altering of a nucleobase of a polynucleotide. 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 cases, the additional domain can facilitate enzymatic or catalytic functions of the base editor, binding functions of the base editor, or be inhibitors of
 15 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 a uracil glycosylase inhibitor (UGI)
 20 domain. A UGI domain can for example improve the efficiency of base editors comprising a cytidine deaminase domain by inhibiting the conversion of a U formed by deamination of a C back to the C nucleobase. In some cases, 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 cases, uracil DNA glycosylase (UDG) can catalyze removal of U from DNA in
 25 cells, which can initiate base excision repair (BER), mostly resulting in reversion of the U:G pair to a C:G pair. In such cases, 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.

30 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, substitution(s) in any domain does/do not change the length of the base editor.

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 cases, 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).

BASE EDITOR SYSTEM

The base editor system provided herein comprises the steps of: (a) contacting a target nucleotide sequence of a polynucleotide (*e.g.*, a double-stranded DNA or RNA, a single-

stranded DNA or RNA) of a subject with a base editor system comprising an adenosine deaminase domain, wherein the aforementioned domains are fused to a polynucleotide binding domain, thereby forming a nucleobase editor capable of inducing changes at one or more bases within a nucleic acid molecule as described herein and at least one guide

5 polynucleic acid (*e.g.*, gRNA), wherein the target nucleotide sequence comprises a targeted nucleobase pair; (b) inducing strand separation of the target region; (c) converting a first nucleobase of the target nucleobase pair in a single strand of the target region to a second nucleobase; and (d) cutting no more than one strand of the target region, where a third nucleobase complementary to the first nucleobase base is replaced by a fourth nucleobase

10 complementary to the second nucleobase. It should be appreciated that in some embodiments, step (b) is omitted. In some embodiments, the 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

15 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

20 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.

Provided herein are systems, compositions, and methods for editing a nucleobase using a base editor system. In some embodiments, the base editor system comprises a base editor (BE) comprising a polynucleotide programmable nucleotide binding domain and a

25 nucleobase editing domain (*e.g.*, deaminase domain) for editing the nucleobase; and a guide polynucleotide (*e.g.*, guide RNA) in conjunction with the polynucleotide programmable nucleotide binding domain. In some embodiments, the base editor system comprises a base editor (BE) comprising a polynucleotide programmable nucleotide binding domain and a nucleobase editing domain (*e.g.*, deaminase domain) for editing the nucleobase, and a guide

30 polynucleotide (*e.g.*, guide RNA) in conjunction with the polynucleotide programmable nucleotide binding domain. 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 cases, a deaminase domain can be an adenine deaminase or an adenosine deaminase. In some embodiments, the terms “adenine deaminase” and “adenosine deaminase” can be used interchangeably. In some cases, a deaminase domain can be an adenine deaminase or an adenosine deaminase. Details of nucleobase editing proteins are described in International PCT Application Nos. 5 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 10 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:eaa04774 (2017), the entire contents of which are hereby incorporated by reference.

15 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 nucleobase components and the polynucleotide programmable nucleotide binding 20 component of a base editor system 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. In some embodiments, a polynucleotide programmable nucleotide binding domain can be fused or linked to a deaminase domain. In some embodiments, a polynucleotide programmable 25 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 can comprise an additional heterologous portion or domain that is capable of interacting with, associating with, or capable of forming a complex with an additional heterologous portion or 30 domain that is part of a polynucleotide programmable nucleotide binding 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. The additional heterologous portion may be a protein domain. In some embodiments, the additional heterologous portion may be a K Homology (KH) domain, a MS2 coat protein domain, a PP7 coat protein domain, a SfMu Com coat protein domain, a sterile alpha motif, a telomerase Ku binding motif and Ku protein, a telomerase Sm7 binding motif and Sm7 protein, or a RNA recognition motif.

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, can comprise 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 (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. The additional heterologous portion may be a protein domain. In some embodiments, the additional heterologous portion may be a K Homology (KH) domain, a MS2 coat protein domain, a PP7 coat protein domain, a SfMu Com coat protein domain, a sterile alpha motif, a telomerase Ku binding motif and Ku protein, a telomerase Sm7 binding motif and Sm7 protein, or a RNA recognition motif.

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
5 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. In some
10 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
15 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 can comprise an additional heterologous portion or domain that is capable of interacting with, associating with, or capable of forming a complex with an additional heterologous portion or domain that is part of a polynucleotide programmable nucleotide
20 binding domain. 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 can comprise 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
25 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
30 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. The additional heterologous portion may be a protein domain. In some embodiments, the additional heterologous portion may be a K Homology (KH) domain, a MS2 coat protein domain, a PP7 coat protein domain, a SfMu Com coat protein domain, a sterile alpha motif, a telomerase Ku binding motif and Ku protein, a telomerase Sm7 binding motif and Sm7 protein, or a RNA recognition motif.

In some embodiments, the base editor inhibits base excision repair 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 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.

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 an 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-XTEN-(SGGS)₂) 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 below Table 9. 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 below Table 9. 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 9 below.

Table 9. 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

	23	26	36	37	48	49	51	72	84	87	106	108	123	125	142	146	147	152	155	156	157	161
ABE5.9	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
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 comprises a monomeric construct containing a TadA*8 variant (“ABE8.x-m”). In some

5 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,

10 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

5 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

15 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

20 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

25 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,

30 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
 5 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
 10 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,
 15 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 10 below.

Table 10: ABE8 base editors

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	Adenosine Deaminase	Adenosine Deaminase Description
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 11

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 11, 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 11: Additional Adenosine Deaminase Base Editor 8 Variants

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.*, ABE9) are generated by cloning an adenosine deaminase variant (*e.g.*, TadA*9) 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, ABE8, or ABE9) is a NGC PAM CP5 variant (*S. pyrogenes* Cas9 or spVRQR Cas9). In some embodiments, the base editor (*e.g.*, ABE7.9, ABE7.10, ABE8, or ABE9) is an AGA PAM CP5 variant (*S. pyrogenes* 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. pyrogenes* 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. pyrogenes* Cas9 or spVRQR Cas9).

In some embodiments, the ABE has a genotype as shown in Table 12 below.

Table 12. 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 13 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 13 below.

Table 13. 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												

	23	36	48	51	76	82	84	106	108	123	146	147	152	154	155	156	157	166
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:

5 ABE8.1_Y147T_CP5_NGC PAM_monomer
MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD **SGGSSGGSSGSETPGTSES**
ATPESSGGSSGGSEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK
10 **GRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD PKKYGGFMQPT**
VAYSVLV VAKVEKGKSKKLKSVKELLGITIMERS SFEKNPIDFLEAKGYKEVKDLIIKLPK
YSLFELENGRKRMLASAKFLQKGNE LALPSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQH
KHYLDEIIIEQISEFSKRVILADANLDKVL SAYNKHRDKPIREQAENIIHLFTLTNLGAPRAF
KYFDTTIARKEYRSTKEVLDATLIHQ SITGLYETRIDLSQLGGDGGSGGSGGSGGSGGSGGS
15 **GGMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHS IKKNLIGALLFDSGETAE**
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVKL
FIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRL ENLIAQLPGEKKNGLFGNLIALSL

In the above sequence, the plain text denotes an adenosine deaminase sequence, bold sequence indicates sequence derived from Cas9, the italics sequence denotes a linker sequence, and the underlined sequence denotes a bipartite nuclear localization sequence.

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:

pNMG-B335 ABE8.1 Y147T CP5 NGC PAM monomer:

251

NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQE
 EFYKFIFKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYF
 LKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASAQSFIER
 MTNFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTN
 5 RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILED
 IVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
 DFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK
 VVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
 QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGK
 10 SDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIT
 KHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLN
 AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEGADKRTADGSEFESPKKKRKV*

In the above sequence, the plain text denotes an adenosine deaminase sequence, bold
 sequence indicates sequence derived from Cas9, the italics sequence denotes a linker
 15 sequence, and the underlined sequence denotes a bipartite nuclear localization sequence.

In some embodiments, the base editor is ABE8.14, which comprises or consists
 essentially of the following sequence or a fragment thereof having adenosine deaminase
 activity:

pNMG-357_ABE8.14 with NGC PAM CP5

20 MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMA
 LRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHP
 GMNHRVEITEGILADECAALLSDFFRMRQEIKAQKKAQSSTDGGSSGGSSGSETPGTSESA
 TPESGGSSGGSSMSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVHNNRVIGEGWNRRAIG
 LHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTG
 25 AAGSLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTDGGSSGG
 SSGSETPGTSESATPESGGSSGGSEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
 TNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDW
 DPKKYGGFMQPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYK
 EVKKDLIIKLPKYSLELENGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPE
 30 DNEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHL
 FTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSI TGLYETRIDLSQLGGDGGSSGG
 GGSSGGSSGGSSGMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLI
 GALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEED
 KKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLLIYLALAHMIKFRGHFLIEG

DLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKK
 NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNL
 SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG
 YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
 5 AILRRQEDFYPLKDNREKIEKILTFRIPYVGPLARGNSRFAWMTRKSEETITPWNFEEV
 DKGASQSFIERMTNFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
 KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDF
 LDNEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLIN
 GIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGS
 10 PAIKKGILQTVKVVDLVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG
 SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDN
 KVLTRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
 IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREI
 NNYHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEGADKRTADGSEF
 15 ESPKKKRKV*

In the above sequence, the plain text denotes an adenosine deaminase sequence, bold sequence indicates sequence derived from Cas9, the italics sequence denotes a linker sequence, and the underlined sequence denotes a bipartite nuclear localization sequence.

In some embodiments, the base editor is ABE8.8-m, which comprises or consists essentially of the following sequence or a fragment thereof having adenosine deaminase activity:

ABE8.8-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHP
 25 GMNHRVEITEGILADECAALLCRFFRMPRVFNAQKKAQSSTDSSGGSSGGSSGSETPGTSES
 ATPESGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
 NPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNG
 30 LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSD
 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
 LRRQEDFYPLKDNREKIEKILTFRIPYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDK
 GASAQSFIERMTNFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK

AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLD
 NEENEDILEDIVLTTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 RDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDLSLHEHIANLAGSPA
 IKKGILQTVKVVDLVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEDIKELGSQ
 5 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNALITQRKFDNLTKAERGGLSELDKAGFIK
 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVINN
 YHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEI GKATAYFFYSNI
 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGG
 10 FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEGKSKKLKSVKELLGI
 TIMERSSEFKNPIDFLEAKGYKEVKDLIIKLPKYSLEFLENGRKRMLASAGELQKGNELAL
 PSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKKHYLDEII EQISEFSKRVLADANLDKV
 LSAYNKHDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSI
 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

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, underlined sequence denotes a bipartite nuclear localization sequence, and double
 underlined sequence indicates mutations.

In some embodiments, the base editor is ABE8.8-d, which comprises or consists
 essentially of the following sequence or a fragment thereof having adenosine deaminase
 activity:

ABE8.8-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMA
 LRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHP
 25 GMNHRVEITEGILADECAALLSDFFRMRQEIKAQKKAQSSTDSSGGSSGGSSGSETPGTSES
 ATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVHNNRVIGEGWNRRAIG
 LHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTG
 AAGSLMDVLHHPGMNHRVEITEGILADECAALLCRRFFRMPRRVFNAQKKAQSSTDSSGGSSGG
 SSGSETPGTSESATPESSGGSSGSSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNT
 30 DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRILIYLAHMI
 KFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLEN
 LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQY
 ADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK

EIFFDQSKNGYAGYIDGGASQEEFYKFIFILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
 PHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEET
 ITPWNFEFVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYEFTVYNELTKVKYVTEGM
 RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHD
 5 LLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTG
 WGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDLSL
 HEHIANLAGSPAIIKKGILQTVKVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMK
 RIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQ
 SFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERG
 10 GLSELDKAGFIKRLVETRQITKHVAQILD SRMNTKYDENDKLIREVKVITLKSCLVSDFRK
 DFQFYKVBREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIASEQEIG
 KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQV
 NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKS
 KKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
 15 AGELQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKHYLDEII EQISEFSK
 RVILADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTK
 EVLDATLIHQ SITGLYETRIDLSQLGGDEGADKRTADGSEFESP KKKRKV*

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
 20 sequence, underlined sequence denotes a bipartite nuclear localization sequence, and double
 underlined sequence indicates mutations.

In some embodiments, the base editor is ABE8.13-m, which comprises or consists
 essentially of the following sequence or a fragment thereof having adenosine deaminase
 activity:

25 ABE8.13-m
 MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLN RVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGLVMQNYRLYDATLYVT FEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHP
 GMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD SGGSSGGSSGSETPGTSES
 ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 30 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHRLEESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIEGDL
 NPDNSDVKLFIQLVQTYNQ LFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKN
 LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLS
 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA

GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRDLLRKQRTFDNGSIPHQIHLGELHAI
 LRRQEDFYFPLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEVVDK
 GASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 5 NEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 RDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDLSLHEHIANLAGSPA
 IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
 10 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVINN
 YHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEI GKATAKYFFYSNI
 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGG
 FSKE SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEGKSKKLKSVKELLGI
 TIMERSSEKPNIDFLEAKGYKEVKKDLIIKLPKYSLENGRKRMLASAGELQKGNELAL
 15 PSKYVNFLYLASHYEKLKGGSPEDNEQQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKV
 LSAYNKHDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI
 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

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
 20 sequence, underlined sequence denotes a bipartite nuclear localization sequence, and double
 underlined sequence indicates mutations.

In some embodiments, the base editor is ABE8.13-d, which comprises or consists
 essentially of the following sequence or a fragment thereof having adenosine deaminase
 activity:

25 ABE8.13-d
 MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMA
 LRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHP
 GMNHRVEITEGILADECAALLSDFFRMRQEIKAQKKAQSSTDSSGGSSGGSSGSETPGTSES
 ATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVHNNRVIGEGWNRRAIG
 30 LHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTG
 AAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRVFNAQKKAQSSTDSSGGSSGG
 SSGSETPGTSESATPESSGGSSGSSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNT
 DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRILIYALAHMI

KFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLN
 LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQY
 ADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK
 EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
 5 PHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPIYYVGPLARGNSRFAMTRKSEET
 ITPWNFEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGM
 RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHD
 LLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTG
 WGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSL
 10 HEHIANLAGSPAIIKKGILQTVKVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMK
 RIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQ
 SFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERG
 GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRK
 DFQFYKVBREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEI
 15 KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQV
 NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKS
 KKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKDLIIKLPKYSLFELNGRKRMLAS
 AGELQKGNELALPSKYVNFYLYLASHYEKLKGSPEDEQQLFVEQHKHYLDEIIEQISEFSK
 RVILADANLDKVL SAYNKH RDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTK
 20 EVLDATLIHQ SITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

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, underlined sequence denotes a bipartite nuclear localization sequence, and double
 underlined sequence indicates mutations.

25 In some embodiments, the base editor is ABE8.17-m, which comprises or consists
 essentially of the following sequence or a fragment thereof having adenosine deaminase
 activity:

ABE8.17-m

MSEVEFSHEYWMRHALLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMA
 30 LRQGGGLVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
 GMNHRVEITEGILADECAALLCYFFRMPRRVFNAQKKAQSSTDSSGSSGGSSGSETPGTSES
 ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRL EESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIEGDL

NPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKNG
 LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSD
 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
 5 LRRQEDFYFPFLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEFVVDK
 GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 RDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
 10 IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEIGIKELGSQ
 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINN
 YHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEI GKATAYFFYSNI
 15 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGG
 FSKE SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEGKSKKLKSVKELLGI
 TIMERSSEKPNIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNELAL
 PSKYVNFLYLASHYEKLKGS PEDNEQQLFVEQHKKHYLDEII EQISEFSKRVILADANLDKV
 LSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI
 20 TGLYETRIDLSQLGGDEGADKRTADGSEFESP KKKRKV*

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, underlined sequence denotes a bipartite nuclear localization sequence, and double underlined sequence indicates mutations.

25 In some embodiments, the base editor is ABE8.17-d, which comprises or consists essentially of the following sequence or a fragment thereof having adenosine deaminase activity:

ABE8.17-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMA
 30 LRQGGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHP
 GMNHRVEITEGILADECAALLSDFFRMRQEIKAQKKAQSSTD SGGSSGGSSGSETPGTSES
 ATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNNRVIGEGWNRRAIG
 LHDPTAHAEIMALRQGGGLVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTG
 AAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRVFNAQKKAQSSTD SGGSSGG

SSGSETPGTSESATPESGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNT
 DRHSIKKNLIGALLEFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSEFFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHMI
 KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEENPINASGVDAKAILSARLSKSRLEN
 5 LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQY
 ADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK
 EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
 PHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPIYYVGPLARGNSRFAMTRKSEET
 ITPWNFEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYEFTVYNELTKVKYVTEGM
 10 RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHD
 LLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTG
 WGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSL
 HEHIANLAGSPAIIKKGILQTVKVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMK
 RIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQ
 15 SFLKDDSIDNKVLTTRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERG
 GLSELDKAGFIKRQLVETRQITKHVAQILD SRMNTKYDENDKLIREVKVITLKSCLVSDFRK
 DFQFYKVVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIG
 KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQV
 NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKS
 20 KKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
 AGELQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQQLFVEQHKHYLDEII EQISEFSK
 RVILADANLDKVLSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTK
 EVLDATLIHQ SITGLYETRIDLSQLGGDEGADKRTADGSEFESP KKKRKV*

In the above sequence, the plain text denotes an adenosine deaminase sequence, bold
 25 sequence indicates sequence derived from Cas9, the italicized sequence denotes a linker
 sequence, underlined sequence denotes a bipartite nuclear localization sequence, and double
 underlined sequence indicates mutations.

In some embodiments, the base editor is ABE8.20-m, which comprises or consists
 essentially of the following sequence or a fragment thereof having adenosine deaminase
 30 activity:

ABE8.20-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLN RVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGLVMQNYRLYDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHP
 GMNHRVEITEGILADECAALLCRRFFRMPRVFNAQKKAQSSTDSSGGSSGGSSGSETPGTSES

ATPESSGGSSGGSDKKYSIGLLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLLIYLALAHMIKFRGHFLIEGDL
 NPDNSDVDFLFIQLVQTYNQLFEEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKN
 5 LFGNLIALSLGLTPNFKS NFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSD
 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIPKILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
 LRRQEDFYFPLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEFVVDK
 GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 10 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTTLTLFEDREMIERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 RDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQDLSLHEHIANLAGSPA
 IKKGILQTVKVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIE EGIKELGSQ
 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSIDNKV
 15 LTRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINN
 YHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNI
 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGG
 FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
 20 TIMERSSEKPNIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL
 PSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKKHYLDEIIIEQISEFSKRVLADANLDKV
 LSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI
 TGLYETRIDLSQLGGDEGADKRTADGSEFESP KKKRKY*

In the above sequence, the plain text denotes an adenosine deaminase sequence, bold
 25 sequence indicates sequence derived from Cas9, the italicized sequence denotes a linker
 sequence, underlined sequence denotes a bipartite nuclear localization sequence, and double
 underlined sequence indicates mutations.

In some embodiments, the base editor is ABE8.20-d, which comprises or consists
 essentially of the following sequence or a fragment thereof having adenosine deaminase
 30 activity:

ABE8.20-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMA
 LRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHP
 GMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD SGGSSGGSSGSETPGTSES

ATPESSGGSSGGSSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIG
 LHDPTAHAEIMALRQGGLVMQNYRLYDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTG
 AAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRVFNAQKKAQSSTDSSGGSSGG
 SSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNT
 5 DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHMI
 KFRGHFLIEGDLNPDNSVDKLEFIQLVQTYNQLFEEENPINASGVDAKAILSARLSKSRLEN
 LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKQLSKDYYDDDLNLLAQIGDQY
 ADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK
 10 EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
 PHQIHLGELHAILRRQEDFYFPFLKDNREKIEKILTFRIPIYVVGPLARGNSRFAMTRKSEET
 ITPWNFEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGM
 RKPFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHD
 LLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTG
 15 WGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDLSL
 HEHIANLAGSPAIIKKGILQTVKVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMK
 RIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQ
 SFLKDDSIDNKVLTTRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERG
 GLSELDKAGFIKRQLVETRQITKHVAQILD SRMNTKYDENDKLIREVKVITLKSCLVSDFRK
 20 DFQFYKVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIG
 KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQV
 NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKS
 KKLKSVKELLGITIMERSSEFKNPIDFLEAKGYKEVKKDIIKLPKYSLFELNGRKRMLAS
 AGELQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQQLFVEQHKHYLDEII EQISEFSK
 25 RVILADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTK
 EVLDATLIHQ SITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

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, underlined sequence denotes a bipartite nuclear localization sequence, and double
 30 underlined sequence indicates mutations.

In some embodiments, an ABE8 is selected from the following sequences:

01. monoABE8.1_bpNLS + Y147T

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP

GMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTDSSGGSSGGSSGSETPGTSES
 ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
 5 NPDNSDVKDLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKN
 LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSD
 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIPHQIHLGELHAI
 LRRQEDFYFPLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDK
 10 GASAQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTTLTFEDREMIERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 RDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
 IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
 15 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSELKAGFIK
 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINN
 YHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNI
 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGG
 20 FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
 TIMERSSEKPNIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASARELQKGNELAL
 PSKYVNFLYLASHYEKLKGSPEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKV
 LSAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDTLIHQSI
 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

25

02. monoABE8.1_bpNLS + Y147R

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
 GMNHRVEITEGILADECAALLCRFFRMPRQVFNAQKKAQSSTDSSGGSSGGSSGSETPGTSES
 30 ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
 NPDNSDVKDLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKN
 LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSD

AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
 LRRQEDFYFPLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDK
 GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 5 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTTLTLFEDREMIERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 RDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
 IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
 10 LTRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINN
 YHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNI
 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIIVKKTEVQTGG
 FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKEGKSKKLKSVKELLGI
 15 TIMERSSEKNPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASARELQKGNELAL
 PSKYVNFLYLASHYEKLKGSPEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKV
 LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSI
 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

20 03. monoABE8.1_bpNLS + Q154S
 MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
 GMNHRVEITEGILADECAALLCYFFRMPRSVFNAQKKAQSSTDSSGGSSGGSSGSETPGTSES
 ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 25 LLFDSGETAEATRLKRTARRRYTRKRNRIQCYLQEIFSNEMAKVDDSFHRLEESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHAMIKFRGHFLIEGDL
 NPDNSDVKDLFIQLVQTYNQLFEENPINASGVDAKILSARLSKSRLENLIAQLPGEKKNL
 LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSD
 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 30 GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
 LRRQEDFYFPLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDK
 GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTTLTLFEDREMIERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI

RDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
 IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
 5 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVVREINN
 YHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNI
 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGG
 FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKEKGKSKKLKSVKELLGI
 TIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELAL
 10 PSKYVNFYLYLASHYEKLKGSPEDEQQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKV
 LSAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDTLIHQSI
 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

04. monoABE8.1_bpNLS + Y123H

15 MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHP
 GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSSGGSSGGSSGSETPGTSES
 ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKK
 20 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
 NPDNSDVKDLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKN
 LFGNLIALSLGLTPNFKSDFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSD
 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLKALVRQQLPEKYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
 25 LRRQEDFYFPLKDNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEVVDK
 GASAQSFIERMTNFDKNLPNEKVLPHKSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTTLTLFEDREMIERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 RDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
 30 IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVVREINN
 YHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNI

MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGG
 FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKEKGKSKKLKSVKELLGI
 TIMERSSEKNPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASARELQKGNELAL
 PSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKV
 5 LSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSI
 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

05. monoABE8.1_bpNLS + V82S

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLV LNNRVIGEGWNRAIGLHDPTAHAEIMA
 10 LRQGGGLVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
 GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSSGSSGGSSGSETPGTSES
 ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHRL EESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIEGDL
 15 NPDNSDVKDLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKN
 LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSD
 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIPILEKMDGTEELLVKNREDLLRKQRTFDNGSIPHQIHLGELHAI
 LRRQEDFY PFLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDK
 20 GASAQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLD
 NEENEDILEDIVLTLT LFEDREMIERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 RDKQSGKTILDFLKSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
 IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
 25 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIK
 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYK VREINN
 YHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMI AKSEQEIGKATAKYFFYSNI
 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGG
 30 FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKEKGKSKKLKSVKELLGI
 TIMERSSEKNPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASARELQKGNELAL
 PSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKV
 LSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSI
 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

06. monoABE8.1_bpNLS + T166R

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
 5 GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSRDSGGSSGGSSGSETPGTSES
 ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRL EESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLLIYLALAHMIKFRGHFLIEGDL
 NPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKN
 10 LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSD
 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
 LRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAMWTRKSEETITPWNFEFVVDK
 GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 15 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 RDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
 IKKGILQTVKVVDLKVVMGRHKPENIVIEARENQTTQKGQKNSRERMKRIE EGIKELGSQ
 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
 20 LTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDAKAGFIK
 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINN
 YHHAHDAYLNAVVGTA LIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI
 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGG
 FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
 25 TIMERSSEKNPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASARELQKGNELAL
 PSKYVNFLYLASHYEKLKGS PEDNEQQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKV
 LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSI
 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

30 07. monoABE8.1_bpNLS + Q154R

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
 GMNHRVEITEGILADECAALLCYFFRMPRRVFNAQKKAQSSTDSSGGSSGGSSGSETPGTSES
 ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDDSFFHRLEESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLLIYLALAHMIKFRGHFLIEGDL
 NPDNSDVKDLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKN
 LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSD
 5 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
 LRRQEDFYFPLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEVVDK
 GASAQSFIERMNTNFDKNLPNEKVLPHKSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLD
 10 NEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 RDKQSGKTILDFLKSDFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
 IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
 15 RQLVETRQITKHVAQILD SRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINN
 YHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNI
 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGG
 FSKE SILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKEKGKSKKLKSVKELLGI
 TIMERS SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASARELQKGNELAL
 20 PSKYVNFYLYLASHYEKLKGS PEDNEQQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV
 LSAYNKH RDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSI
 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

08. monoABE8.1_bpNLS + Y147R_Q154R_Y123H

25 MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHP
 GMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDSSGSSGGSSGSETPGTSES
 ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDDSFFHRLEESFLVEEDKK
 30 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLLIYLALAHMIKFRGHFLIEGDL
 NPDNSDVKDLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKN
 LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSD
 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI

LRRQEDFY PFLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEVVDK
 GASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTLTTLFEDREMIEERLKTIAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 5 RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
 IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
 RQLVETRQITKHVAQILD SRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYKVREINN
 10 YHHAHDAYLNAVVG TALIKKYPKLESEFVYG DYKVYDVRKMI AKSEQEIGKATAKYFFYSNI
 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGG
 FSKESILPKRNSDKLIARKKDWD PKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
 TIMERSSEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGKRMLASARELQKGNELAL
 PSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKKHYLDEIIEQISEFSKRVLADANLDKV
 15 LSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSI
 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

09. monoABE8.1_bpNLS + Y147R_Q154R_I76Y

MSEVEFSHEYWMRHALLAKRARDEREVPVGAVLV LNNRVIGEGWNRAIGLHDPTAHAEIMA
 20 LRQGGLVMQNYRLYDATLYVT FEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
 GMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDSSGSSGGSSGSETPGTSES
 ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHRL EESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIEGDL
 25 NPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKN
 LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSD
 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
 LRRQEDFY PFLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEVVDK
 30 GASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTLTTLFEDREMIEERLKTIAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
 IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ

ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDAKAGFIK
 RQLVETRQITKHVAQILD SRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYK VREINN
 YHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI
 5 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGG
 FSKESILPKRNSDKLIARKKDWD PKKYGGFVSPTVAYSVLVVAKEGKSKKLKSVKELLGI
 TIMERSSEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELAL
 PSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKV
 LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL DATLIHQSI
 10 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

10. monoABE8.1_bpNLS + Y147R_Q154R_T166R

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLV LNNRVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGLVMQNYRLIDATLYVT FEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
 15 GMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSRDSGGSSGGSSGSETPGTSES
 ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHRL EESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLR LIYLALAHMIKFRGHFLIEGDL
 NPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKN
 20 LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADLFLAAKNLSD
 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
 LRRQEDFYFPFLKDNREKIEKILTFRIPIYVGPLARGNSRFAMWTRKSEETITPWNFE EVVDK
 GASAQSFIERM TNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 25 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTLT LFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 RDKQSGKTILDFLKS DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
 IKKGILQTVKVVD ELVKVMGRHKPENIV IEMARENQTTQKGQKNSRERMKRIE EGIKELGSQ
 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
 30 LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDAKAGFIK
 RQLVETRQITKHVAQILD SRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYK VREINN
 YHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI
 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGG
 FSKESILPKRNSDKLIARKKDWD PKKYGGFVSPTVAYSVLVVAKEGKSKKLKSVKELLGI

TIMERSSEKPNIDFLEAKGYKEVKKDLIIKLPKYSLEFLENGRKRMLASARELQKGNELAL
 PSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKV
 LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL DATLIHQSI
 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

5

11. monoABE8.1_bpNLS + Y147T_Q154R

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
 GMNHRVEITEGILADECAALLCTFFRMPRRVFNAQKKAQSSTDSSGSSGGSSGSETPGTSES

10

ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRL EESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIEGDL
 NPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNG
 LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSD

15

AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIPHQIHLGELHAI
 LRRQEDFYFPLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEVVVDK
 GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD

20

NEENEDILEDIVLTTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 RDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
 IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK

25

RQLVETRQITKHVAQILD SRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYK VREINN
 YHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMI AKSEQEIGKATAKYFFYSNI
 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGG
 FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
 TIMERSSEKPNIDFLEAKGYKEVKKDLIIKLPKYSLEFLENGRKRMLASARELQKGNELAL

30

PSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKV
 LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL DATLIHQSI
 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

12. monoABE8.1_bpNLS + Y147T_Q154S

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2
CONTENANT LES PAGES 1 À 270

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 2
CONTAINING PAGES 1 TO 270

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

What is claimed is:

1. An adenosine deaminase comprising an alteration at an amino acid position selected from the group consisting of 21, 23, 25, 38, 51, 54, 70, 71, 72, 73, 94, 124, 133, 139, 146, and 158 of SEQ ID NO: 1, or a corresponding alteration in another adenosine deaminase:

5 10 20 30 40 50
MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG
 60 70 80 90 100
LHDPTAHAEI MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG
10 110 120 130 140 150
RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCYFFR
 160
MPRQVFNAAQK KAQSSTD (SEQ ID NO: 1).

15 2. The adenosine deaminase of claim 1, which comprises an alteration selected from the group consisting of R21N, R23H, E25F, N38G, L51W, P54C, M70V, Q71M, N72K, Y73S, M94V, P124W, T133K, D139L, D139M, C146R, and A158K of SEQ ID NO: 1, or a corresponding alteration in another adenosine deaminase.

20 3. The adenosine deaminase of claim 1 or 2, further comprising a V82T alteration of SEQ ID NO: 1, or a corresponding alteration in another adenosine deaminase.

4. The adenosine deaminase of any one of claims 1-3, which comprises alterations at two or more amino acid positions selected from the group consisting of 21, 23, 25, 38, 51, 54,
25 70, 71, 72, 73, 94, 124, 133, 139, 146, and 158 of SEQ ID NO: 1, or a corresponding alteration in another adenosine deaminase.

5. The adenosine deaminase of any one of claims 1-4, which comprises two or more of said alterations.

30

6. The adenosine deaminase of any one of claims 1-5, which comprises three or more of said alterations.

7. The adenosine deaminase of any one of claims 1-6, which further comprises one or
35 more of the following alterations: Y147T, Y147R, Q154S, Y123H, and Q154R.

8. The adenosine deaminase of any one of claims 1-7, wherein the adenosine deaminase comprises any one of the following groups of alterations:

E25F + V82S + Y123H;

T133K + Y147R + Q154R;

5 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;

10 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;

15 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;

20 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;

25 M70V + V82S + M94V + Y123H + Y147R + Q154R;

Q71M + V82S + Y123H + Y147R + Q154R;

E25F + I76Y + V82S + Y123H + Y147R + Q154R;

I76Y + V82T + Y123H + Y147R + Q154R;

N38G + I76Y + V82S + Y123H + Y147R + Q154R;

30 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;
 5 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;
 10 V82S + Q154R;
 N72K + V82S + Y123H + Y147R + Q154R;
 Q71M + V82S + Y123H + Y147R + Q154R;
 V82S + Y123H + T133K + Y147R + Q154R;
 V82S + Y123H + T133K + Y147R + Q154R + A158K;
 15 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;
 20 V82S + Y123H + T133K + Y147R + Q154R + A158K; or
 M70V + Q71M + N72K + V82S + Y123H + Y147R + Q154R.

9. The adenosine deaminase of any one of claims 1-8, which comprises a deletion of the
 C terminus beginning at a residue selected from the group consisting of 149, 150, 151, 152,
 25 153, 154, 155, 156, and 157.

10. The adenosine deaminase of any one of claims 1-6, further comprising an alteration
 selected from the group consisting of Y147T, Y147R, Q154S, Y123H, V82S, T166R, and
 Q154R.

30

11. The adenosine deaminase of any one of claims 1-6, which is an adenosine deaminase
 variant described in Table 14, Table 18, or FIGS. 3A-3C.

12. A fusion protein comprising a polynucleotide programmable DNA binding domain and at least one base editor domain that is an adenosine deaminase variant comprising an alteration at an amino acid position selected from the group consisting of 21, 23, 25, 38, 51, 54, 70, 71, 72, 73, 94, 124, 133, 139, 146, and 158 of SEQ ID NO: 1, or a corresponding alteration in another adenosine deaminase:

10

10	20	30	40	50
MSEVEFSHEY	WMRHALTLAK	<u>R</u> <u>A</u> <u>R</u> <u>D</u> <u>E</u> REVPV	GAVLVLN <u>N</u> RV	IGEGWNRAIG
60	70	80	90	100
<u>L</u> <u>H</u> <u>D</u> <u>P</u> TAHA EI	MALRQGGLV <u>M</u>	<u>Q</u> <u>N</u> <u>Y</u> RLIDATL	YVTFEPCVMC	AGAM <u>I</u> HSRIG
110	120	130	140	150
RVVFGVRNAK	TGAAGSLMDV	LHY <u>P</u> GMNHRV	EI <u>T</u> EGILA <u>D</u> E	CAALL <u>C</u> YFFR
160				
MPRQVFNA	<u>A</u> QK	KAQSSTD	(SEQ ID NO: 1).	

15 13. The fusion protein of claim 12, wherein the adenosine deaminase variant comprises an alteration selected from the group consisting of R21N, R23H, E25F, N38G, L51W, P54C, M70V, Q71M, N72K, Y73S, M94V, P124W, T133K, D139L, D139M, C146R, and A158K of SEQ ID NO: 1, or a corresponding alteration in another adenosine deaminase.

20 14. A fusion protein comprising a polynucleotide programmable DNA binding domain and at least one base editor domain that is an adenosine deaminase variant comprising an alteration selected from the group consisting of R21N, R23H, E25F, N38G, L51W, P54C, M70V, Q71M, N72K, Y73S, M94V, P124W, T133K, D139L, D139M, C146R, and A158K of SEQ ID NO: 1, or a corresponding alteration in another adenosine deaminase.

25

15. The fusion protein of any one of claims 12-14, further wherein the adenosine deaminase variant further comprises a V82T alteration of SEQ ID NO: 1, or a corresponding alteration in another adenosine deaminase.

30 16. A fusion protein comprising a polynucleotide programmable DNA binding domain and at least one base editor domain that is an adenosine deaminase variant comprising an alteration V82T and one or more alterations selected from the group consisting of R21N, R23H, E25F, N38G, L51W, P54C, M70V, Q71M, N72K, Y73S, M94V, P124W, T133K, D139L, D139M, C146R, and A158K of SEQ ID NO: 1, or a corresponding alteration in
35 another adenosine deaminase.

17. The fusion protein of any one of claims 12-16, wherein the adenosine deaminase variant comprises alterations at two or more amino acid positions selected from the group consisting of 21, 23, 25, 38, 51, 54, 70, 71, 72, 73, 94, 124, 133, 139, 146, and 158 of SEQ ID NO: 1, or a corresponding alteration in another adenosine deaminase.

5

18. The fusion protein of any one of claims 12-17, wherein the adenosine deaminase variant comprises two or more of said alterations.

10

19. The fusion protein of any one of claims 12-17, wherein the adenosine deaminase variant comprises three or more of said alterations.

20. The fusion protein of any one of claims 12-19, wherein the adenosine deaminase variant further comprises one or more of the following alterations: Y147T, Y147R, Q154S, Y123H, and Q154R.

15

21. The fusion protein of any one of claims 12-20, wherein the adenosine deaminase variant comprises any one of the following groups of alterations:

E25F + V82S + Y123H;

T133K + Y147R + Q154R;

20 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;

25 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;

30 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;
 5 N72K + V82S + Y123H + D139L + Y147R + Q154R;
 E25F + V82S + Y123H + D139M + Y147R + Q154R;
 M70V + V82S + M94V + Y123H + Y147R + Q154R;
 Q71M + V82S + Y123H + Y147R + Q154R;
 E25F + I76Y + V82S + Y123H + Y147R + Q154R;
 10 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;
 15 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;
 20 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;
 25 V82S + Q154R;
 N72K + V82S + Y123H + Y147R + Q154R;
 Q71M + V82S + Y123H + Y147R + Q154R;
 V82S + Y123H + T133K + Y147R + Q154R;
 V82S + Y123H + T133K + Y147R + Q154R + A158K;
 30 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; or

M70V + Q71M + N72K + V82S + Y123H + Y147R + Q154R.

- 5 22. The fusion protein of any one of claims 12-20, wherein the adenosine deaminase variant comprises a deletion of the C terminus beginning at a residue selected from the group consisting of 149, 150, 151, 152, 153, 154, 155, 156, and 157.

- 10 23. The fusion protein of any one of claims 12-20, wherein the base editor domain comprises an adenosine deaminase variant monomer, wherein the adenosine deaminase monomer comprises one or more alterations selected from the group consisting of R21N, R23H, E25F, N38G, L51W, P54C, M70V, Q71M, N72K, Y73S, V82T, M94V, P124W, T133K, D139L, D139M, C146R, and A158K of SEQ ID NO: 1.

- 15 24. The fusion protein of any one of claims 12-17, wherein the base editor domain comprises an adenosine deaminase heterodimer comprising a wild-type adenosine deaminase domain and an adenosine deaminase variant.

- 20 25. The fusion protein of claim 24, where the adenosine deaminase variant further comprises an alteration selected from the group consisting of Y147T, Y147R, Q154S, Y123H, V82S, T166R, and Q154R.

- 25 26. The fusion protein of any one of claims 12-17, wherein the base editor domain comprises an adenosine deaminase heterodimer comprising a TadA*7.10 domain and adenosine deaminase variant domain.

27. The fusion protein of claim 26, where the adenosine deaminase variant comprises two or more alterations.

- 30 28. The fusion protein of any one of claims 12-17, wherein the base editor comprises a heterodimer comprising a TadA*7.10 domain and an adenosine deaminase variant comprising any one of the following groups of alterations:
E25F + V82S + Y123H;

- T133K + Y147R + Q154R;
 E25F + V82S + Y123H + Y147R + Q154R;
 L51W + V82S + Y123H + C146R + Y147R + Q154R;
 Y73S + V82S + Y123H + Y147R + Q154R;
 5 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;
 10 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;
 15 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;
 20 N72K + V82S + Y123H + D139L + Y147R + Q154R;
 E25F + V82S + Y123H + D139M + Y147R + Q154R;
 M70V + V82S + M94V + Y123H + Y147R + Q154R;
 Q71M + V82S + Y123H + Y147R + Q154R;
 E25F + I76Y + V82S + Y123H + Y147R + Q154R;
 25 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;
 30 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;
 5 I76Y + V82S + Y123H + D139M + Y147R + Q154R;
 Y73S + I76Y + V82S + Y123H + Y147R + Q154R;
 V82S + Q154R;
 N72K + V82S + Y123H + Y147R + Q154R;
 Q71M + V82S + Y123H + Y147R + Q154R;
 10 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;
 15 M70V + V82S + M94V + Y123H + Y147R + Q154R;
 V82S + Y123H + T133K + Y147R + Q154R;
 V82S + Y123H + T133K + Y147R + Q154R + A158K; or
 M70V + Q71M + N72K + V82S + Y123H + Y147R + Q154R.
- 20 29. The fusion protein of any one of claims 12-17, wherein the adenosine deaminase variant is an ABE9 or TadA*9 deaminase variant described in Table 14, Table 18, or FIGS. 3A-3C.
- 25 30. The fusion protein of any one of claims 12-29, wherein the adenosine deaminase variant is a truncated ABE8 or ABE9 that 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 ABE9.
- 30 31. The fusion protein of any one of claims 12-30, wherein the polynucleotide programmable DNA binding domain is a Cas9, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ domain.
32. A fusion protein comprising a polynucleotide programmable DNA binding domain comprising the following sequence:

EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
 VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPPKKYGGFM
 QPTVAYSVLVVAKEKGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKE
 VKKDLIIKLPKYSLFELNGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYE
 5 KLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVLSAYNKHR
 DKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSITE
 LYETRIDLSQLGGDGGSGGSGGSGGSGGSGGSGGMDKKYSIGLAIGTNSVGWAVI
 TDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR
 KNRICYLQEIFSNEMAKVDDSSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
 10 KYPTIYHLRKKLVDSTDKADLRILIYALAHMIKFRGHFLIEGDLNPDNSDVKL
 FIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFG
 NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADFLAA
 KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPPEKYK
 EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
 15 RTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPIYYVGPLARG
 NSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
 LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK
 EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKHKDKDFLDNEENEDILEDIVL
 TLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQ
 20 SGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA
 GSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERM
 KRIEIEGKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSD
 YDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLN
 AKLITQRKFDNLTKAERGGLSELKAGFIKRQLVETRQITKHVAQILDSRMNTK
 25 YDENDKLIREVKVITLKSCLVSDFRKDFQFYKREINNYHHAHDAYLNAVVGTA
 LIKKYPKLESEFVYGDYKVYDVRKMIKSEQEGADKRTADGSEFESPKKKRKV*,

wherein the bold sequence indicates sequence derived from Cas9, the italics sequence
 denotes a linker sequence, and the underlined sequence denotes a bipartite nuclear
 localization sequence, and at least one base editor domain comprising an adenosine
 30 deaminase variant comprising an alteration at an amino acid position selected from the group
 consisting of 21, 23, 25, 38, 51, 54, 70, 71, 72, 73, 94, 124, 133, 138, 139, 146, and 158 of
 SEQ ID NO: 1.

33. The fusion protein of claim 32, wherein the adenosine deaminase variant comprises an alteration selected from the group consisting of R21N, R23H, E25F, N38G, L51W, P54C, M70V, Q71M, N72K, Y73S, M94V, P124W, T133K, D138M, D139L, D139M, C146R, and A158K of SEQ ID NO: 1.

5

34. The fusion protein of claim 33, wherein the adenosine deaminase variant comprises an alteration V82T of SEQ ID NO: 1.

10

35. The fusion protein of claim 33 or 34, wherein the adenosine deaminase variant comprises two or more of said alterations.

36. The fusion protein of claim 33 or 34, wherein the adenosine deaminase variant comprises three or more of said alterations.

15

37. The fusion protein of claim 33 or 34, wherein the adenosine deaminase variant further comprises an alteration selected from the group consisting of Y147T, Y147R, Q154S, Y123H, V82S, T166R, and Q154R.

20

38. The fusion protein of claim 33 or 34, wherein the adenosine deaminase variant comprises two or more of the following alterations: Y147T, Y147R, Q154S, Y123H, and Q154R.

25

39. The fusion protein of claim 32, wherein the adenosine deaminase variant comprises any one of the following groups of alterations:

E25F + V82S + Y123H;

T133K + Y147R + Q154R;

E25F + V82S + Y123H + Y147R + Q154R;

L51W + V82S + Y123H + C146R + Y147R + Q154R;

Y73S + V82S + Y123H + Y147R + Q154R;

30

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;
- 5 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;
- 10 R21N + V82S + Y123H + Y147R + Q154R;
 V82S + Y123H + Y147R + Q154R + A158K;
 N72K + V82S + Y123H + D139L + Y147R + Q154R;
 E25F + V82S + Y123H + D139M + Y147R + Q154R;
 M70V + V82S + M94V + Y123H + Y147R + Q154R;
- 15 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;
- 20 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;
- 25 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;
- 30 I76Y + V82S + Y123H + D139M + Y147R + Q154R;
 Y73S + I76Y + V82S + Y123H + Y147R + Q154R;
 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;
 5 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;
 10 M70V + Q71M + N72K + V82S + Y123H + Y147R + Q154R; or
 or any other alteration or group thereof of Table 14.

40. The fusion protein of any one of claims 12-39, wherein the polynucleotide
 programmable DNA binding domain is a *Staphylococcus aureus* Cas9 (SaCas9),
 15 *Streptococcus thermophilus* 1 Cas9 (St1Cas9), a *Streptococcus pyogenes* Cas9 (SpCas9), or
 variants thereof.

41. The fusion protein of any one of claims 12-40, wherein the polynucleotide
 programmable DNA binding domain comprises a modified SaCas9 having an altered
 20 protospacer-adjacent motif (PAM) specificity.

42. The fusion protein of claim 41, wherein the modified SaCas9 comprises amino acid
 substitutions E782K, N968K, and R1015H, or a corresponding amino acid substitutions
 thereof.

25

43. The fusion protein of any one of claims 12-40, wherein the polynucleotide
 programmable DNA binding domain comprises a variant of SpCas9 having an altered
 protospacer-adjacent motif (PAM) specificity.

- 30 44. The fusion protein of claim 43, wherein the altered PAM has specificity for the
 nucleic acid sequence 5'-NGA-3', 5'-NGC-3', 5'-NGG-3', 5'-NGT-3', or 5'-NGN-3'.

45. The fusion protein of claim 43 or 44, wherein the variant SpCas9 comprises amino acid substitutions selected from:

D1135M, S1136Q, G1218K, E1219F, A1322R, D1332A, R1335E, and T1337R, or corresponding amino acid substitutions thereof;

5 I322V, S409I, E427G, R654L, R753G (MQKFRAER) or corresponding amino acid substitutions thereof;

I322V, S409I, E427G, R654L, R753G, R1114G or corresponding amino acid substitutions thereof; or amino acid substitutions as set forth in FIGS. 3A-3C.

10 46. The fusion protein of any one of claims 12-45, wherein the polynucleotide programmable DNA binding domain is a nuclease inactive or nickase variant.

47. The fusion protein of claim 46, wherein the nickase variant comprises an amino acid substitution D10A or a corresponding amino acid substitution thereof.

15

48. The fusion protein of any one of claims 12-47, wherein the adenosine deaminase domain is capable of deaminating adenine in deoxyribonucleic acid (DNA).

20 49. The fusion protein of any one of claims 12-47, wherein the adenosine deaminase is a modified adenosine deaminase that does not occur in nature.

50. The fusion protein of any one of claims 12-49, wherein the adenosine deaminase is a TadA deaminase.

25 51. The fusion protein of claim 50, wherein the TadA deaminase is a TadA*7.10 variant.

52. The fusion protein of any one of claims 12-51, comprising a linker between the polynucleotide programmable DNA binding domain and the adenosine deaminase domain.

30 53. The fusion protein of claim 52, wherein the linker comprises the amino acid sequence: SGGSSGGSSGSETPGTSESATPES.

54. The fusion protein of any one of claims 12-53, comprising one or more nuclear localization signals.

55. The fusion protein of claim 54, wherein the nuclear localization signal is a bipartite nuclear localization signal.

56. The fusion protein of any one of claims 12-55, wherein the Cas9 is a StCas9.

57. The fusion protein of any one of claims 12-55, wherein the Cas9 is a SaCas9 or an SpCas9.

58. The fusion protein of any one of claims 12-55, wherein the Cas9 is a modified SaCas9 or a modified SpCas9.

59. The fusion protein of claim 58, wherein the modified SaCas9 comprises amino acid substitutions E782K, N968K, and R1015H, or corresponding amino acid substitutions thereof.

60. The fusion protein of claim 59, wherein the modified SaCas9 comprises the amino acid sequence:

KRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRR
RRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRR
GVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKT
SDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEW
YEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYEYEFQIIEN
VFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIENA
ELLDQIAKILTIYQSSEDIQEELTNLNSLTQEEIEQISNLKGYTGTHNLSLKAINLILDE
LWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIK
KYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIK
LHDMQEGKCLYSLEAIPLEDLLNPNFNYEVDHIIPRSVSFDNSFNKVLVKQEENSKK
GNRTPFYQLSSSDSKISYETFKKHILNLA KGKGRISKTKKEYLLEERDINRFSVQKDFI
NRNLVDTRYATRGLMNLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKG
YKHAEDALIIANADFIFKEWKLDKAKKVMENQMFEKQAESMPEIETE QEYKEIF
ITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNNLNGLYDK

DNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTK
 YSKKDNGPVIKKIKYYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVY
 KFTVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYKNDLIKINGELYS
 VIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLY
 5 EVKSKKHPQIIKKG.

61. A polynucleotide encoding the fusion protein of any one of claims 12-60.
62. A cell produced by introducing into the cell, or a progenitor thereof:
 10 a polynucleotide encoding the fusion protein of any one of claims 12-60, and
 one or more guide polynucleotides that target the base editor to effect an A•T to G•C
 alteration of a SNP associated with a genetic disease.
63. The cell of claim 62, wherein the cell is a human cell.
 15
64. The cell of claim 62 or 63, wherein the cell is *in vitro* or *in vivo*.
65. The cell of any one of claims 62-64, wherein the genetic disease is alpha-1 antitrypsin
 deficiency (A1AD).
 20
66. The cell of any one of claims 62-65, wherein the fusion protein and the one or more
 guide polynucleotides forms a complex in the cell.
67. An isolated cell or population of cells propagated or expanded from the cell of any
 25 one of claims 62-66.
68. A method of treating a genetic disease in a subject in need thereof, the method
 comprising administering to the subject a cell of any one of claims 62-67.
69. The method of claim 68, wherein the cell is autologous, allogeneic, or xenogeneic to
 30 the subject.

70. A base editor system comprising a polynucleotide programmable DNA binding domain and at least one base editor domain that is an adenosine deaminase variant comprising an alteration at an amino acid position selected from the group consisting of 21, 23, 25, 38, 51, 54, 70, 71, 72, 73, 82, 94, 124, 133, 139, 146, and 158 of SEQ ID NO: 1, or a corresponding alteration in another adenosine deaminase:

10

	10	20	30	40	50
	MSEVEFSHEY	WMRHALTLAK	<u>R</u> <u>A</u> <u>R</u> <u>D</u> <u>E</u> REVPV	GAVLVLN <u>N</u> RV	IGEGWNRAIG
	60	70	80	90	100
	<u>L</u> H <u>D</u> <u>P</u> TAHA EI	MALRQGGLV <u>M</u>	<u>Q</u> <u>N</u> <u>Y</u> RLIDATL	YVTFEPCVMC	AGAM <u>I</u> HSRIG
10	110	120	130	140	150
	RVVFGVRNAK	TGAAGSLMDV	LHY <u>P</u> GMNHRV	EI <u>T</u> EGILA <u>D</u> E	CAALL <u>C</u> YFFR
	160				
	MPRQVFNA <u>A</u> QK	KAQSSTD	(SEQ ID NO: 1).		

15 71. The base editor system of claim 70, wherein the adenosine deaminase variant comprises an alteration selected from the group consisting of R21N, R23H, E25F, N38G, L51W, P54C, M70V, Q71M, N72K, Y73S, V82T, M94V, P124W, T133K, D139L, D139M, C146R, and A158K of SEQ ID NO: 1, or a corresponding alteration in another adenosine deaminase.

20

72. The base editor system of claim 70 or 71, further comprising one or more guide polynucleotides that target the base editor domain to effect an A•T to G•C alteration of a SNP associated with a genetic disease.

25 73. The base editor system of any one of claims 70-72, wherein the adenosine deaminase variant is capable of deaminating adenine in deoxyribonucleic acid (DNA).

74. The base editor system of claim 73, wherein the guide polynucleotide comprises ribonucleic acid (RNA), or deoxyribonucleic acid (DNA).

30

75. The base editor system of claim 74, wherein the guide polynucleotide comprises a CRISPR RNA (crRNA) sequence, a trans-activating CRISPR RNA (tracrRNA) sequence, or a combination thereof.

35 76. The base editor system of claim 72, further comprising a second guide polynucleotide.

77. The base editor system of claim 76, wherein the second guide polynucleotide comprises ribonucleic acid (RNA), or deoxyribonucleic acid (DNA).

78. The base editor system of claim 76, wherein the second guide polynucleotide
5 comprises a CRISPR RNA (crRNA) sequence, a trans-activating CRISPR RNA (tracrRNA) sequence, or a combination thereof.

79. The base editor system of any one of claims 70-78, wherein the polynucleotide-programmable DNA-binding domain comprises a Cas9, Cas12a/Cpf1, Cas12b/C2c1,
10 Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ domain.

80. The base editor system of claim 79, wherein the polynucleotide-programmable DNA-binding domain is nuclease dead.

15

81. The base editor system of claim 79, wherein the polynucleotide-programmable DNA-binding domain is a nickase.

82. The base editor system of claim 79, wherein the polynucleotide-programmable DNA-binding domain comprises a Cas9 domain.

20

83. The base editor system of claim 82, wherein the Cas9 domain comprises a nuclease dead Cas9 (dCas9), a Cas9 nickase (nCas9), or a nuclease active Cas9.

84. The base editor system of claim 83, wherein the Cas9 domain comprises a Cas9 nickase.

25

85. The base editor system of any one of claims 70-84, wherein the polynucleotide-programmable DNA-binding domain is an engineered or a modified polynucleotide-programmable DNA-binding domain.

30

86. The base editor system of claim 72, wherein the genetic disease is alpha-1 antitrypsin deficiency (A1AD).

87. A method for correcting a single nucleotide polymorphism (SNP) in a polynucleotide: contacting a target nucleotide sequence, at least a portion of which is located in the polynucleotide or its reverse complement, with a fusion protein of any one of claims 12-60 or the base editor system of any one of claims 70-85; and editing the SNP by deaminating the SNP or its complement nucleobase upon targeting of the base editor to the target nucleotide sequence, wherein deaminating the SNP or its complement nucleobase corrects the SNP.

88. The method of claim 87, wherein the SNP is associated with alpha-1 antitrypsin deficiency (A1AD).

89. The method of claim 87 or 88, wherein the SNP is in the *SERPINA1* gene and the correction comprises an E342K (PiZ allele) alteration.

90. A method for editing a polynucleotide, the method comprising contacting a target nucleotide sequence with the fusion protein of any one of claims 12-60 or the base editor system of any one of claims 70-85, thereby editing the polynucleotide.

91. The method of claim 90, wherein the editing results in less than 20% indel formation, less than 15% indel formation, less than 10% indel formation; less than 5% indel formation; less than 4% indel formation; less than 3% indel formation; less than 2% indel formation; less than 1% indel formation; less than 0.5% indel formation; or less than 0.1% indel formation.

92. The method of claim 91, wherein the editing does not result in translocations.

93. A base editor comprising an ABE9 comprising a TadA*7.10 adenosine deaminase variant domain and a Cas9 endonuclease domain selected from the following:

monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+A109S of SEQ ID NO: 1, and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER);
monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+T111R of SEQ ID NO: 1, and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER);
monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+D119N of SEQ ID NO: 1, and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER);

monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+H122N of SEQ ID NO: 1, and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER);

monoTadA*7.10 having mutations I76Y+V82T+Y147d+Q154S of SEQ ID NO: 1, and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER);

5 monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+F149Y of SEQ ID NO: 1, and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER);

monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+T166I of SEQ ID NO: 1, and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER); and

10 monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+D167N of SEQ ID NO: 1, and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER).

mono TadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+L36H+N157K of SEQ ID NO: 1, and spCas9 having mutations I322V, S409I, E427G, R654L, R753G, R1114G (MQKFRAER);

15 mono TadA*7.10 having mutations I76Y+V82T+Y147D+Q154S+F149Y+D167N+L36H+N157K of SEQ ID NO: 1, and SpCas9 having mutations I322V, S409I, E427G, R654L, R753G, R1114G (MQKFRAER);

monoTadA*7.10 having mutations I76Y+V82T+Y147D+Q154S+F149Y+D167N+L36H+N157K+V106W of SEQ ID NO: 1, and SpCas9 having mutations I322V, S409I, E427G, R654L, R753G, R1114G (MQKFRAER);

mono TadA*7.10 having mutations A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N of SEQ ID NO: 1, and SpCas9 having mutations I322V, S409I, E427G, R654L, R753G, R1114G, MQKFRAER; and

25 mono TadA*7.10 having mutations A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N + V106W of SEQ ID NO: 1, and SpCas9 having mutations I322V, S409I, E427G, R654L, R753G, R1114G (MQKFRAER); and one or more guide polynucleotides that target the adenosine deaminase variant domain to effect an A•T to G•C alteration of a SNP associated with a genetic disease.

30

94. The base editor of claim 93, wherein the SNP is associated with alpha-1 antitrypsin deficiency (A1AD).

95. A vector comprising one or more polynucleotides encoding an ABE9 base editor comprising a TadA adenosine deaminase domain and an SpCas9 endonuclease domain selected from

- monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+A109S and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER);
- monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+T111R and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER);
- monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+D119N and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER);
- 10 monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+H122N and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER);
- monoTadA*7.10 having mutations I76Y+V82T+Y147d+Q154S and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER);
- monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+F149Y and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER);
- 15 monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+T166I and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER); and
- monoTadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+D167N and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER).
- 20 mono TadA*7.10 having mutations I76Y+V82T+Y147T+Q154S+L36H+N157K and spCas9 having mutations I322V, S409I, E427G, R654L, R753G, R1114G (MQKFRAER);
- mono TadA*7.10 having mutations I76Y+V82T+Y147D+Q154S+F149Y+D167N+L36H+N157K and SpCas9 having mutations I322V, S409I, E427G, R654L, R753G, R1114G (MQKFRAER);
- 25 monoTadA*7.10 having mutations I76Y+V82T+Y147D+Q154S+F149Y+D167N+L36H+N157K+V106W and SpCas9 having mutations I322V, S409I, E427G, R654L, R753G, R1114G, (MQKFRAER)
- mono TadA*7.10 having mutations A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N and SpCas9 having mutations I322V, S409I, E427G, R654L, R753G, R1114G (MQKFRAER); and
- 30 mono TadA*7.10 having mutations A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N + V106W and SpCas9 having mutations I322V, S409I, E427G, R654L, R753G, R1114G (MQKFRAER).

96. The vector of claim 95, which is a plasmid, viral, or mRNA vector.

97. A composition comprising the fusion protein of any one of claims 12-60 or the base
5 editor system of any one of claims 70-85.

98. The composition of claim 97, further comprising a pharmaceutically acceptable
excipient, diluent, or carrier.

10 99. A composition comprising the fusion protein of any one of claims 12-60 bound to a
guide RNA, wherein the guide RNA comprises a nucleic acid sequence that is
complementary to an *SERPINA1* gene associated with alpha-1 antitrypsin deficiency
(A1AD).

15 100. A composition comprising the base editor system of any one of claims 70-85 bound to
a guide RNA, wherein the guide RNA comprises a nucleic acid sequence that is
complementary to an *SERPINA1* gene associated with alpha-1 antitrypsin deficiency
(A1AD).

20 101. The composition of any one of claims 97-100, wherein the adenosine deaminase
variant is capable of deaminating adenine in deoxyribonucleic acid (DNA).

102. The composition of any one of claims 97-101, wherein the fusion protein or base
editor system

25 (i) comprises a Cas9 nickase;

(ii) comprises a nuclease inactive Cas9;

(iii) comprises an SpCas9 variant comprising a combination of amino acid
substitutions shown in FIGS. 3A-3C; or

(iv) comprises an SpCas9 variant comprising a combination of amino acid sequence
30 substitutions selected from I322V, S409I, E427G, R654L, R753G (MQKFRAER); or I322V,
S409I, E427G, R654L, R753G, R1114G, (MQKFRAER).

103. The composition of any one of claims 99-102, further comprising a pharmaceutically acceptable excipient, diluent, or carrier.

104. A pharmaceutical composition for the treatment of a disease or disorder comprising
5 the composition of claim 98.

105. The pharmaceutical composition of claim 104, wherein the disease or disorder is alpha-1 antitrypsin deficiency (A1AD).

106. The pharmaceutical composition of claim 105, wherein the fusion protein or the base editor system is bound to a guide RNA, wherein the guide RNA comprises a nucleic acid sequence that is complementary to an *SERPINA1* gene associated with alpha-1 antitrypsin deficiency (A1AD).

107. The pharmaceutical composition of claim 106, wherein the gRNA and the base editor are formulated together or separately.

108. The pharmaceutical composition of any one of claims 98, or 103-107, wherein the gRNA comprises a nucleic acid sequence, from 5' to 3', or a 1, 2, 3, 4, or 5 nucleotide 5' truncation fragment thereof, selected from one or more of

5'-ACCAUCGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC
AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU
CGGUGCUUUU-3';

5'-CCAUCGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC
25 AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU
CGGUGCUUUU-3';

5'-CAUCGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU
AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3';

5'-AUCGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU
30 AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3';

5'-UCGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU
AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3'; or

5'-CGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU
AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3'.

109. The pharmaceutical composition of any one of claims 98, or 103-108, further
5 comprising a vector suitable for expression in a mammalian cell, wherein the vector
comprises a polynucleotide encoding the base editor.

110. The pharmaceutical composition of claim 109, wherein the polynucleotide encoding
the base editor is mRNA.

10

111. The pharmaceutical composition of claim 109, wherein the vector is a viral vector.

112. The pharmaceutical composition of claim 111, wherein the viral vector is a retroviral
vector, adenoviral vector, lentiviral vector, herpesvirus vector, or adeno-associated viral
15 vector (AAV).

113. The pharmaceutical composition of any one of claims 98, or 103-108, further
comprising a ribonucleoparticle suitable for expression in a mammalian cell.

20

114. The pharmaceutical composition of any one of claims 98, or 103-108, further
comprising a lipid.

115. A method of treating alpha-1 antitrypsin deficiency (A1AD), the method comprising
administering to a subject in need thereof the pharmaceutical composition of any one of
25 claims 98 or 103-114.

116. Use of the pharmaceutical composition of any one of claims 98 or 103-114 in the
treatment of alpha-1 antitrypsin deficiency (A1AD) in a subject.

30

117. The method of claim 115 or the use of claim 116, wherein the subject is a human.

118. The base editor system of any one of claims 70-86, wherein the adenosine deaminase
variant comprises any one of the following groups of alterations:

- E25F + V82S + Y123H;
 T133K + Y147R + Q154R;
 E25F + V82S + Y123H + Y147R + Q154R;
 L51W + V82S + Y123H + C146R + Y147R + Q154R;
 5 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;
 10 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;
 15 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;
 20 V82S + Y123H + Y147R + Q154R + A158K;
 N72K + V82S + Y123H + D139L + Y147R + Q154R;
 E25F + V82S + Y123H + D139M + Y147R + Q154R;
 M70V + V82S + M94V + Y123H + Y147R + Q154R;
 Q71M + V82S + Y123H + Y147R + Q154R;
 25 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;
 30 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;

5 R21N + I76Y + V82S + Y123H + Y147R + Q154R;

I76Y + V82S + Y123H + D139M + Y147R + Q154R;

Y73S + I76Y + V82S + Y123H + Y147R + Q154R;

V82S + Q154R;

N72K + V82S + Y123H + Y147R + Q154R;

10 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;

15 Q71M + V82S + Y123H + Y147R + Q154R;

M70V + V82S + M94V + Y123H + Y147R + Q154R;

V82S + Y123H + T133K + Y147R + Q154R;

V82S + Y123H + T133K + Y147R + Q154R + A158K; or

M70V + Q71M + N72K + V82S + Y123H + Y147R + Q154R.

20

119. The adenosine deaminase, fusion protein, base editor, or base editor system of any one of the preceding claims, wherein the adenosine deaminase or adenosine deaminase variant is a TadA*7.10 variant comprising any one of the following amino acid alterations or groups of alterations:

25

V82T;

I76Y + V82T; or

I76Y + V82T + Y147T + Q154S.

30

120. An adenosine deaminase variant which is a TadA*7.10 variant comprising any one of the following amino acid alterations or groups of alterations:

V82T;

I76Y + V82T; or

I76Y + V82T + Y147T + Q154S.

121. A fusion protein comprising a polynucleotide programmable DNA binding domain and at least one base editor domain that is an TadA*7.10 adenosine deaminase variant comprising any one of the following amino acid alterations or groups of alterations:

- 5 V82T;
 I76Y + V82T; or
 I76Y + V82T + Y147T + Q154S.

122. The fusion protein of claim 121, wherein the polynucleotide programmable DNA
10 binding domain comprises a Cas9 endonuclease domain.

123. The fusion protein of claim 122, wherein the Cas9 endonuclease domain comprises spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER).

15 124. The adenosine deaminase variant of claim 121 or the fusion protein of any one of claims 121-123, wherein the TadA7*10 is monomeric.

125. A nucleobase editor comprising a TadA*7.10 adenosine deaminase variant domain and a Cas9 endonuclease domain selected from the following:

- 20 monoTadA*7.10 having mutation V82T and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER);
 monoTadA*7.10 having mutations I76Y + V82T and spCas9 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER); or
 monoTadA*7.10 having mutations I76Y + V82T + Y147T + Q154S and spCas9
25 having mutations I322V, S409I, E427G, R654L, R753G (MQKFRAER).

1/11

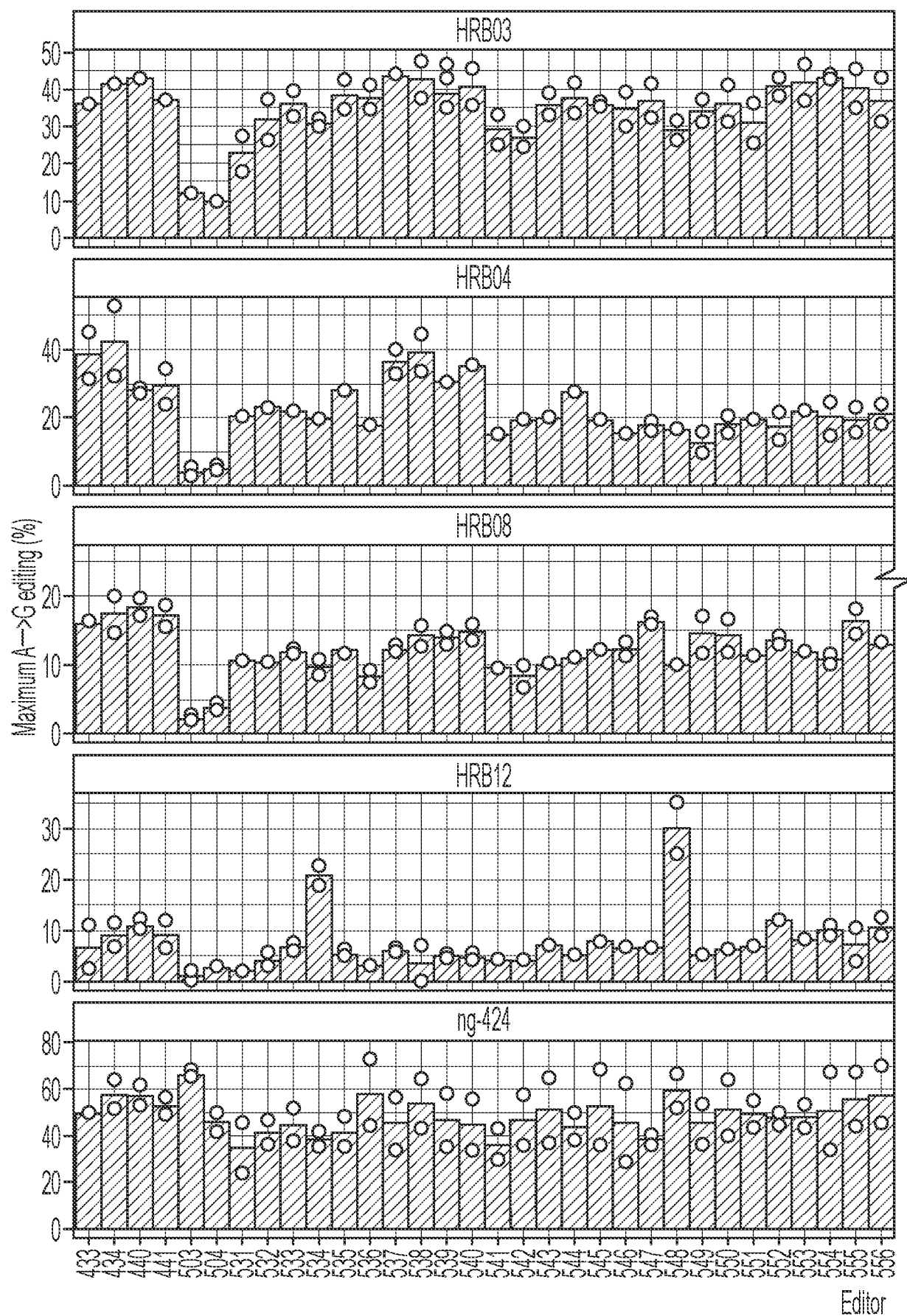


FIG. 1

2/11

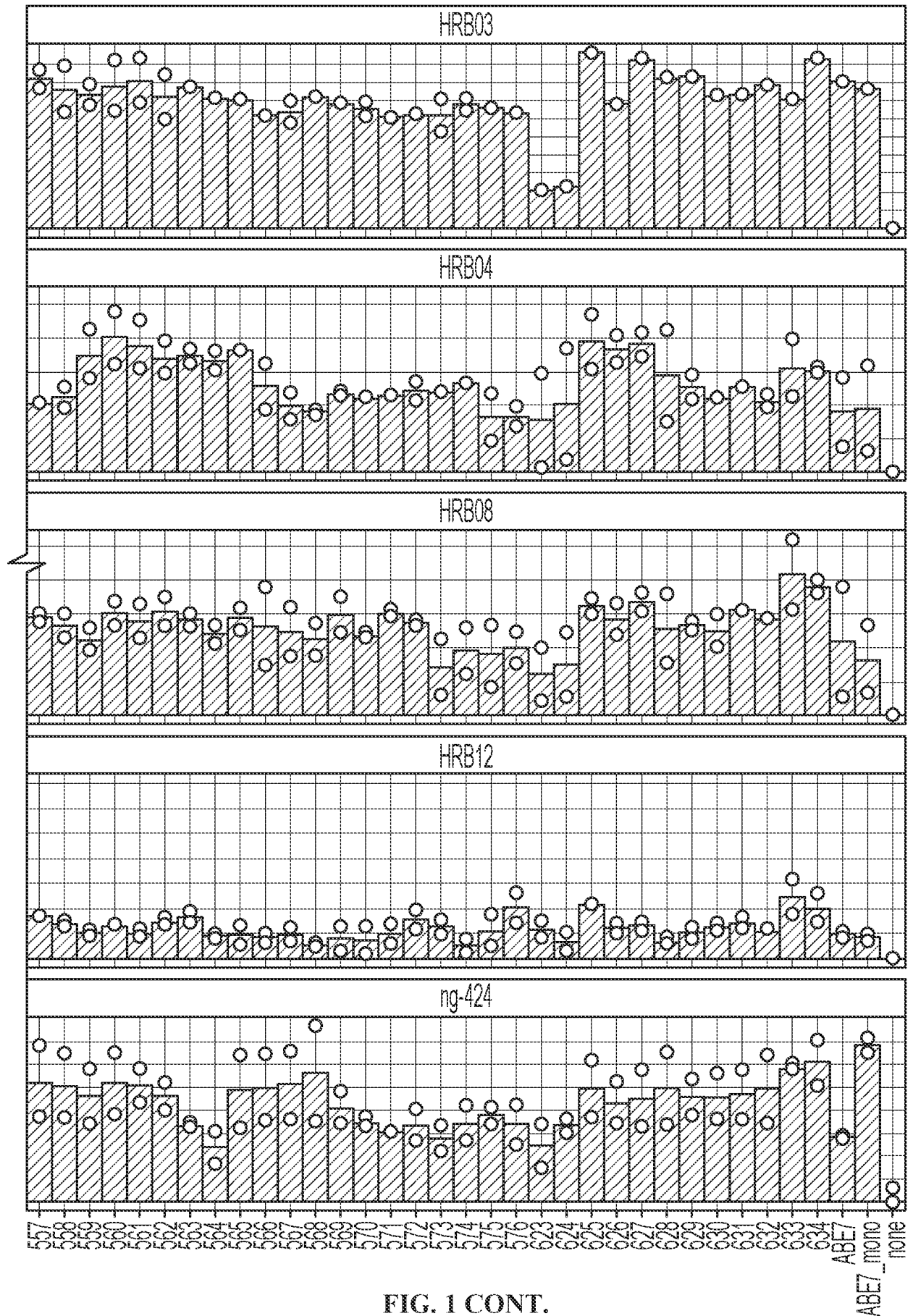


FIG. 1 CONT.

3/11

	HRB03			HRB04		HRB08		
none	0.08%	0.02%	0.0033%	0.024%	0.01%	0.013%	0.0089%	0.0078%
ABE7_mono	39%	31%	0.96%	17%	9.2%	7.5%	5.2%	0.28%
ABE7	40%	26%	1%	17%	7.1%	10%	5.6%	1.3%
634	47%	22%	1.7%	30%	24%	19%	11%	1%
633	35%	26%	0.73%	31%	26%	21%	14%	1.8%
632	39%	23%	0.71%	21%	14%	14%	9.5%	1%
631	37%	27%	0.88%	25%	22%	16%	11%	1.1%
630	37%	16%	0.4%	22%	15%	13%	8.9%	0.76%
629	42%	16%	1.2%	25%	15%	13%	7.1%	0.64%
628	41%	25%	1.1%	29%	15%	13%	6%	0.84%
627	46%	34%	1.4%	38%	21%	17%	7.8%	1.6%
626	34%	23%	1.1%	37%	18%	14%	6.5%	0.92%
625	48%	25%	1.4%	39%	21%	16%	7.7%	1.3%
624	11%	1.7%	0.44%	19%	11%	7.6%	4.5%	0.8%
623	11%	1%	0.02%	15%	9.4%	5%	4.4%	0.18%
576	32%	20%	0.45%	16%	14%	10%	7.6%	0.6%
575	33%	22%	0.58%	16%	14%	8.9%	6.3%	0.77%
574	34%	24%	0.66%	26%	17%	9.5%	5.6%	0.45%
573	31%	19%	0.64%	24%	14%	7.1%	4.5%	0.46%
572	32%	21%	0.44%	24%	17%	14%	7.4%	1.2%
571	30%	24%	0.82%	23%	19%	15%	10%	1.2%
570	33%	21%	0.6%	22%	16%	12%	7.5%	0.66%
569	34%	24%	0.93%	23%	19%	15%	9.2%	0.98%
568	36%	23%	0.6%	18%	14%	11%	7%	0.79%
567	32%	17%	0.48%	20%	12%	12%	7.1%	0.94%
566	31%	15%	0.58%	26%	13%	13%	5.1%	1.1%
565	35%	26%	0.98%	37%	20%	14%	7.1%	0.88%
564	36%	24%	0.63%	33%	17%	12%	6.9%	1.3%
563	39%	24%	1.2%	34%	19%	14%	6.7%	1.6%
562	36%	28%	1%	34%	19%	15%	7.9%	0.67%
561	40%	29%	0.98%	38%	20%	14%	7.6%	0.97%
560	39%	28%	1%	40%	24%	15%	8%	1.3%
559	37%	15%	0.56%	35%	14%	11%	6%	0.57%
558	38%	25%	0.86%	22%	20%	13%	9.4%	1%
557	41%	31%	0.84%	18%	21%	14%	11%	1%

FIG. 2

4/11

HRB12			ng-424				
0.019%	0.049%	0.019%	0.007%	0.55%	0.34%	0.32%	0.009%
1.6%	4.5%	2.9%	4.9%	68%	41%	10%	0.63%
3.4%	4.5%	2.7%	2.3%	28%	15%	3.3%	1.2%
2.9%	10%	4.7%	4.5%	61%	49%	12%	2.1%
3.8%	12%	6.1%	8.3%	59%	47%	16%	1.9%
1.6%	5.7%	2.9%	2.6%	49%	34%	11%	0.98%
2.6%	7%	4%	4.1%	47%	34%	13%	1.4%
1.7%	6.3%	2.4%	2.7%	46%	36%	9%	0.94%
1.7%	5.2%	1.8%	2.6%	46%	33%	9.1%	0.62%
2.6%	3.4%	0.85%	4.8%	49%	33%	8.3%	0.31%
2.7%	6.5%	3.4%	7.5%	45%	35%	11%	1.4%
2.6%	6.1%	2.1%	6.8%	43%	33%	8.6%	1%
3.6%	11%	4.1%	7%	49%	38%	11%	2%
1.2%	3.2%	1.5%	3.4%	33%	20%	5.5%	0.69%
1.2%	5.9%	1.2%	3.8%	24%	18%	6.8%	0.6%
2.4%	10%	4.2%	2.1%	34%	25%	7.7%	0.67%
2%	5.6%	1.8%	3.2%	38%	28%	8.1%	0.41%
1.5%	2.4%	0.5%	2.3%	34%	25%	6.9%	0.89%
1.1%	6.3%	3.1%	2.6%	28%	21%	8%	0.76%
1.1%	7.7%	2.9%	1.8%	33%	26%	7.9%	1.1%
0.92%	4.8%	2.8%	3.2%	31%	23%	9.4%	0.68%
0.86%	3.7%	1.4%	2%	35%	27%	8.2%	0.92%
0.82%	3.8%	2%	4%	41%	32%	15%	2%
1.6%	2.8%	1.7%	1.5%	56%	43%	15%	2%
2%	4.6%	0.98%	4.3%	52%	42%	18%	2.4%
2%	4.1%	0.88%	3.5%	50%	28%	6%	0.32%
1.9%	4.5%	2.9%	4.5%	48%	33%	6%	0.5%
1.7%	4.4%	1.8%	3.5%	23%	17%	5.9%	0.38%
4.4%	8.3%	4%	4.4%	34%	25%	8.2%	1.3%
3.9%	7.3%	3.2%	9.2%	46%	37%	14%	1.2%
2.3%	5%	2.2%	6.6%	51%	39%	11%	1.2%
3.6%	6.6%	2.4%	9.4%	52%	38%	15%	1.6%
1.6%	5.1%	1.4%	6%	46%	34%	8.6%	0.85%
2.4%	7%	4.1%	2.8%	51%	38%	11%	1.3%
2.6%	8.6%	5.4%	2.5%	52%	42%	12%	1.3%

% A→G

60

40

20

FIG. 2 CONT.

5/11

556	37%	15%	0.89%	21%	17%	13%	9%	0.48%
555	40%	23%	0.92%	19%	18%	16%	12%	0.92%
554	43%	33%	0.86%	20%	17%	11%	6.4%	1.1%
553	42%	31%	0.92%	22%	22%	12%	7%	0.82%
552	41%	21%	1.1%	18%	14%	14%	7.8%	1.1%
551	31%	22%	0.64%	20%	19%	12%	9.3%	1%
550	36%	24%	0.76%	15%	18%	14%	8.7%	1%
549	34%	22%	0.62%	13%	12%	14%	8.8%	0.99%
548	29%	6.4%	2.6%	13%	17%	8.7%	9.9%	0.64%
547	37%	23%	0.8%	18%	16%	16%	9.7%	0.84%
546	35%	17%	0.57%	16%	13%	12%	9.4%	0.56%
545	36%	16%	0.66%	19%	17%	12%	9.8%	0.74%
544	38%	19%	0.72%	27%	13%	11%	6.6%	1%
543	36%	14%	0.41%	20%	15%	10%	7.1%	0.55%
542	27%	5.7%	0.24%	19%	8.3%	8.3%	5.2%	0.16%
541	29%	13%	0.39%	15%	9.1%	9.6%	6.1%	0.13%
540	41%	28%	0.55%	36%	22%	15%	7.9%	1.3%
539	39%	24%	0.8%	30%	19%	14%	7.8%	1%
538	43%	15%	0.79%	39%	18%	14%	6.2%	0.91%
537	44%	28%	0.76%	36%	21%	12%	8.4%	0.72%
536	38%	10%	0.53%	18%	12%	8.2%	4.4%	0.41%
535	38%	20%	0.55%	28%	14%	12%	6.1%	0.95%
534	31%	4%	1.1%	20%	11%	9.7%	7.7%	0.53%
533	36%	13%	0.74%	22%	12%	12%	7.5%	1.4%
532	32%	8.7%	0.51%	23%	10%	10%	5.4%	0.64%
531	23%	5.3%	0.28%	20%	8%	11%	5.6%	0.6%
504	10%	1.6%	0.08%	3.3%	4.4%	1.6%	3.8%	0.16%
503	12%	1.7%	0.46%	3.4%	3.4%	1.7%	2.2%	0.18%
441	37%	25%	0.85%	27%	26%	17%	13%	1.5%
440	43%	29%	1.4%	28%	23%	18%	12%	1.6%
434	41%	29%	0.72%	42%	27%	17%	9.4%	1.3%
433	36%	26%	1.1%	38%	21%	16%	8.7%	1.2%
	7	10	11	3	9	3	8	12

Target A Position

FIG. 2 CONT.

6/11

4.5%	11%	2.2%	5.4%	58%	47%	12%	0.64%
3.6%	7.2%	2.2%	3.8%	35%	44%	9.8%	0.76%
3.7%	10%	4.2%	2.4%	51%	36%	12%	1.4%
3.1%	8.4%	2.2%	1.9%	48%	37%	8.9%	1.7%
3.1%	12%	6.3%	3.4%	48%	33%	13%	3.1%
2.5%	6.9%	3.6%	3%	49%	39%	12%	2%
1.3%	5.9%	4.4%	1.2%	52%	34%	12%	0.46%
0.65%	5.2%	1.6%	2.1%	45%	34%	8.4%	0.62%
9.9%	30%	9.7%	8.6%	59%	40%	20%	0.7%
1.9%	6.5%	3%	2.4%	38%	30%	7.2%	1.4%
1.6%	5.7%	4.4%	2.9%	46%	33%	9.2%	1%
2.9%	7.8%	3.2%	3.2%	53%	39%	12%	0.82%
1.2%	5.2%	2.7%	1.6%	43%	33%	7.4%	0.88%
2.6%	7.1%	2.1%	1.5%	51%	36%	8.7%	0.53%
1.5%	4.3%	0.58%	2.5%	47%	31%	4.9%	0.14%
1.4%	4.3%	0.37%	1.6%	36%	22%	3.9%	0.13%
1.5%	4.9%	3.4%	3.5%	45%	33%	8.8%	1%
1.5%	5%	3%	3%	47%	34%	7%	0.46%
0.56%	3.6%	3.5%	3.3%	54%	38%	7.9%	1.4%
1.4%	6%	2.2%	4.2%	45%	32%	8.3%	0.26%
2%	3.2%	1.2%	2.8%	58%	43%	12%	0.24%
1.3%	5.5%	2.9%	2.6%	41%	32%	5.8%	0.42%
7.1%	21%	4.9%	1.7%	39%	14%	5.3%	1%
2.4%	6.9%	2.8%	0.84%	45%	34%	7.9%	2%
1.2%	4.1%	0.96%	1.7%	41%	29%	5.4%	0.62%
1.2%	2.1%	0.64%	1.1%	34%	22%	3.5%	0.13%
0.02%	2.8%	0.02%	0.18%	46%	22%	2.7%	0.49%
0.26%	1.1%	0.22%	0.3%	67%	25%	5.8%	0.48%
1.9%	9.2%	4.6%	3.7%	53%	44%	14%	1.2%
2.8%	11%	3.6%	4%	57%	47%	15%	0.93%
3.2%	9%	4.6%	8.3%	58%	45%	15%	1.1%
3.4%	6.6%	2.6%	3.3%	50%	40%	10%	1.1%
2	8	11	2	5	7	8	11

FIG. 2 CONT.

FIG. 3A

MSP785	A109S	pMRNA-T7-5'Trlink-monoTadA-ABE7.10(I76Y+V82T+Y147T+Q154S+A109S)-spCas9(I322V,S409I,E427G,R654L,R753G_MQKFRAER) 120A BbsI
MSP786	T111R	pMRNA-T7-5'Trlink-monoTadA-ABE7.10(I76Y+V82T+Y147T+Q154S+T111R)-spCas9(I322V,S409I,E427G,R654L,R753G_MQKFRAER) 120A BbsI
MSP787	D119N	pMRNA-T7-5'Trlink-monoTadA-ABE7.10(I76Y+V82T+Y147T+Q154S+D119N)-spCas9(I322V,S409I,E427G,R654L,R753G_MQKFRAER) 120A BbsI
MSP788	H122N	pMRNA-T7-5'Trlink-monoTadA-ABE7.10(I76Y+V82T+Y147T+Q154S+H122N)-spCas9(I322V,S409I,E427G,R654L,R753G_MQKFRAER) 120A BbsI
MSP789	Y147d	pMRNA-T7-5'Trlink-monoTadA-ABE7.10(I76Y+V82T+Y147d+Q154S)-spCas9(I322V,S409I,E427G,R654L,R753G_MQKFRAER) 120A BbsI
MSP790	F149Y	pMRNA-T7-5'Trlink-monoTadA-ABE7.10(I76Y+V82T+Y147T+Q154S+F149Y)-spCas9(I322V,S409I,E427G,R654L,R753G_MQKFRAER) 120A BbsI
MSP791	T166I	pMRNA-T7-5'Trlink-monoTadA-ABE7.10(I76Y+V82T+Y147T+Q154S+T166I)-spCas9(I322V,S409I,E427G,R654L,R753G_MQKFRAER) 120A BbsI
MSP792	D167N	pMRNA-T7-5'Trlink-monoTadA-ABE7.10(I76Y+V82T+Y147T+Q154S+D167N)-spCas9(I322V,S409I,E427G,R654L,R753G_MQKFRAER) 120A BbsI

7/11

FIG. 3B

Name	TadA	TadA variant	PAM variant
ngcABEvar9	Mono	TadA_7.10+H76Y+V82T+Y147T+Q154S	MQKFRAER
ngcABEvar10	Mono	TadA_7.10+I76Y+V82T+Y147T+Q154S	I322V,S409I,E427G,R654L,R753G, MQKFRAER
ngcABEvar11	Mono	TadA_7.10+I76Y+V82T+Y147T+Q154S	I322V,S409I,E427G,R654L,R753G, R1114G, MQKFRAER
ngcABEvar12	Mono	TadA_7.10+H76Y+V82T+Y147T+Q154S+L36H+N157K	I322V,S409I,E427G,R654L,R753G, R1114G, MQKFRAER
ngcABEvar13	Mono	TadA_7.10+H76Y+V82T+Y147d+Q154S+F149Y+D167N+L36H+N157K	I322V,S409I,E427G,R654L,R753G, R1114G, MQKFRAER
ngcABEvar14	Mono	ABE8e	I322V,S409I,E427G,R654L,R753G, R1114G, MQKFRAER

8/11

FIG. 3C

TadA	TadA variant	PAM variant	120A template
Mono	TadA*7.10+I76Y+V82T+Y147T+Q154S+L36H+N157K	I322V, S409I, E427G, R654L, R753G, R1114G, MQKFRAER	MSP784
Mono	TadA*7.10+I76Y+V82T+Y147D+Q154S+F149Y+D167N+L36H+N157K	I322V, S409I, E427G, R654L, R753G, R1114G, MQKFRAER	MSP819
Mono	ABE8e (Monomer_TadA*7.10 + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N)	I322V, S409I, E427G, R654L, R753G, R1114G, MQKFRAER	MSP820
Mono	TadA*7.10+I76Y+V82T+Y147D+Q154S+F149Y+D167N+L36H+N157K+V106W	I322V, S409I, E427G, R654L, R753G, R1114G, MQKFRAER	MSP821
Mono	ABE8e (Monomer_TadA*7.10 + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N)+V106W	I322V, S409I, E427G, R654L, R753G, R1114G, MQKFRAER	MSP822

9/11

FIG. 4A



FIG. 4B

Code name	TadA	TadA variant	PAM variant
ngcABEvar1	Dual	TadA 7.10	MQKSER
ngcABEvar2	Dual	TadA 7.10	RVRFRAR
ngcABEvar3	Dual	TadA 7.10	MQKFRAER
ngcABEvar4	Dual	TadA 7.10+V82S	MQKFRAER
ngcABEvar5	Mono	TadA 7.10+V82S	MQKFRAER
ngcABEvar6	Mono	TadA 7.10+I76Y+V82S	MQKFRAER
ngcABEvar7	Mono	TadA 7.10+V82T	MQKFRAER
ngcABEvar8	Mono	TadA 7.10+I76Y+V82T	MQKFRAER
ngcABEvar9	Mono	TadA 7.10+I76Y+V82T +Y147T+Q154S	MQKFRAER

FIG. 4D

Variant 9 Outperform Variant 8

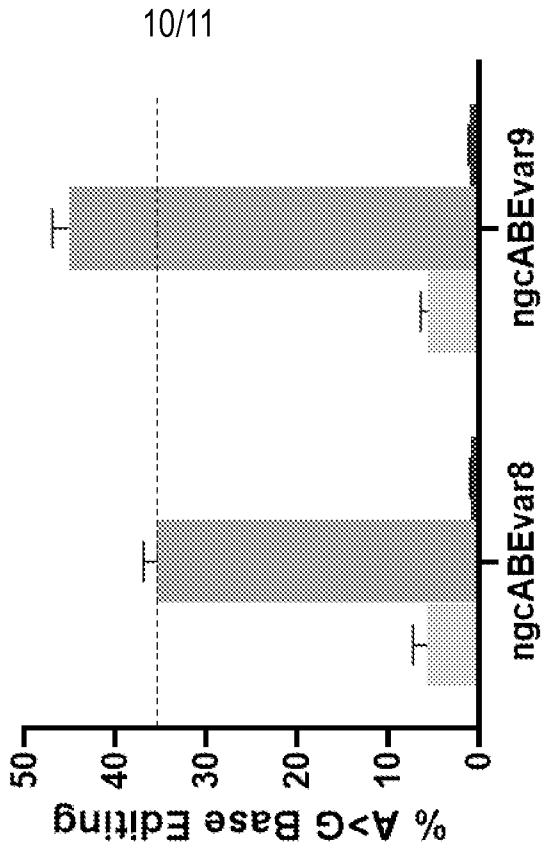
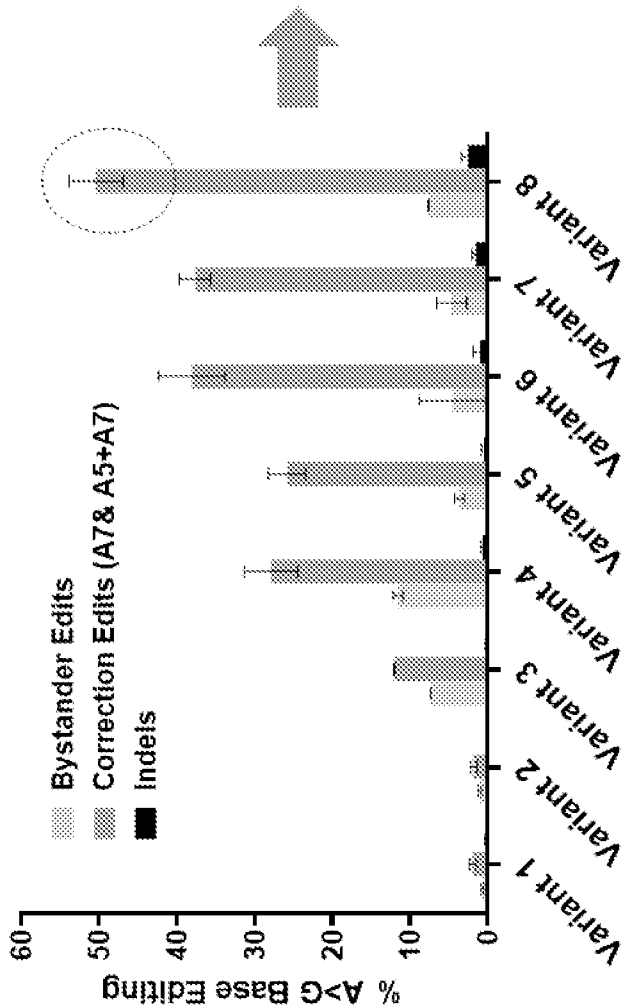


FIG. 4C

E342K Correction in PIZ Fibroblasts



11/11

FIG. 5A

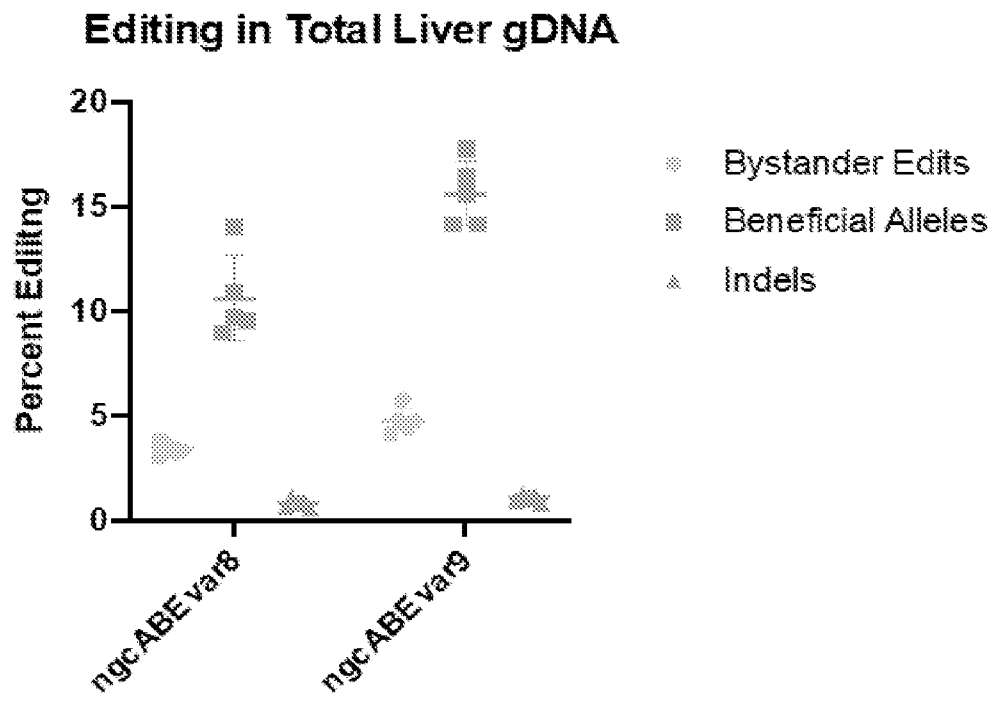


FIG. 5B

