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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 November 2023 (16.11.2023)

WIPO | PCT

(10) International Publication Number
WO 2023/220654 A2

(51) International Patent Classification:
Not classified

(21) International Application Number:
PCT/US2023/066848

(22) International Filing Date:
10 May 2023 (10.05.2023)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

63/340,377	10 May 2022 (10.05.2022)	US
63/340,433	10 May 2022 (10.05.2022)	US
63/351,714	13 June 2022 (13.06.2022)	US
63/353,977	21 June 2022 (21.06.2022)	US
63/380,933	25 October 2022 (25.10.2022)	US
63/383,845	15 November 2022 (15.11.2022)	US
63/483,907	08 February 2023 (08.02.2023)	US

(71) Applicant: **MAMMOTH BIOSCIENCES, INC.** [US/US]; 1000 Marina Blvd, Suite 600, Brisbane, California 94005 (US).

(72) Inventors: **CHEN, Sean**; c/o Mammoth Biosciences, Inc., 1000 Marina Blvd, Suite 600, Brisbane, California 94005 (US). **HARRINGTON, Lucas Benjamin**; c/o Mammoth Biosciences, Inc., 1000 Marina Blvd, Suite 600, Brisbane, California 94005 (US). **WRIGHT, William Douglass**; c/o Mammoth Biosciences, Inc., 1000 Marina Blvd, Suite 600, Brisbane, California 94005 (US). **DELOUGHERY, Aaron**; c/o Mammoth Biosciences, Inc., 1000 Marina Blvd, Suite 600, Brisbane, California 94005 (US). **HARTONO, Wiputra Jaya**; c/o Mammoth Biosciences, Inc., 1000 Marina Blvd, Suite 600, Brisbane, California 94005 (US). **RAUCH, Benjamin Julius**; c/o Mammoth Biosciences, Inc., 1000 Marina Blvd, Suite 600, Brisbane, California 94005 (US). **TYMOSHENKO, Stepan**; c/o Mammoth Biosciences, Inc., 1000 Marina Blvd, Suite 600, Brisbane, California 94005 (US).

(74) Agent: **HEBERT, Micheal L.** et al.; c/o Greenberg Traurig LLP, 77 West Wacker Drive, Suite 3100, Chicago, Illinois 60601 (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,

(54) Title: EFFECTOR PROTEIN COMPOSITIONS AND METHODS OF USE THEREOF

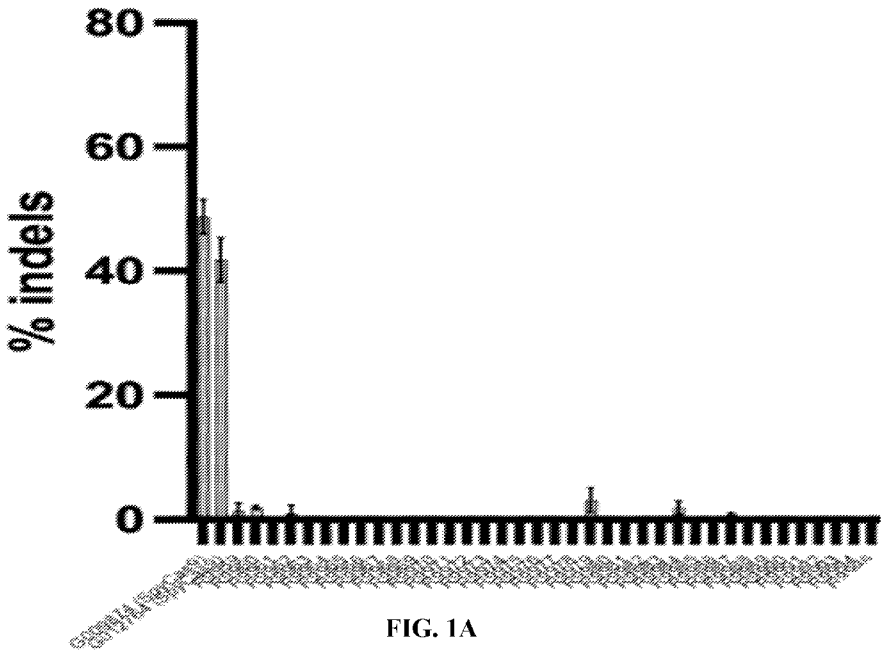

FIG. 1A

(57) Abstract: Provided herein are compositions, systems, and methods comprising effector proteins and uses thereof. These effector proteins may be characterized as CRISPR-associated (Cas) proteins. Various compositions, systems, and methods of the present disclosure may leverage the activities of these effector proteins for the modification, and engineering of nucleic acids.

WO 2023/220654 A2

[Continued on next page]

CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

— *with sequence listing part of description (Rule 5.2(a))*

EFFECTOR PROTEIN COMPOSITIONS AND METHODS OF USE THEREOF

CROSS-REFERENCE

[1] This application claims benefit of U.S. Provisional Application No. 63/340,433, filed on May 10, 2022, 63/351,714, filed on June 13, 2022, 63/353,977, filed on June 21, 2022, 63/380,933, filed on October 25, 2022, 63/483,907, filed on February 8, 2023, 63/340,377, filed on May 10, 2022, 63/383,845, filed on November 15, 2022, which is incorporated herein by reference in its entirety.

INCORPORATION BY REFERENCE OF SEQUENCE LISTING

[2] The instant application contains a Sequence Listing, which has been submitted via Patent Center. The Sequence Listing titled "203477-757601-PCT.xml", which was created on May 09, 2023 and is 1,879,767 bytes in size, is hereby incorporated by reference in its entirety.

FIELD

[3] The present disclosure relates generally to compositions of effector proteins and guide nucleic acids, and methods and systems of using such compositions, including expressing a functional human protein, as well as, the treatment of disorders associated with irregular expression of the functional human protein, for instance, monogenic diseases.

BACKGROUND

[4] Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and associated proteins (Cas proteins), sometimes referred to as a CRISPR/Cas system, were first identified in certain bacterial species and are now understood to form part of a prokaryotic acquired immune system. CRISPR/Cas systems provide immunity in bacteria and archaea against viruses and plasmids by targeting the nucleic acids of the viruses and plasmids in a sequence-specific manner. While CRISPR/Cas proteins are involved in the acquisition, targeting and cleavage of foreign DNA or RNA, the systems may also contain a CRISPR array, which includes direct repeats flanking short spacer sequences that, in part, guide Cas proteins to their targets. The discovery of CRISPR/Cas systems has revolutionized the field of genomic manipulation and engineering, and therapeutic applications of these systems are being explored.

SUMMARY

[5] The present disclosure provides for compositions, systems, and methods comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, and uses thereof. Compositions, systems, and methods disclosed herein leverage nucleic acid modifying activities of these effector proteins and guide nucleic acids for inserting a gene encoding a functional human protein into a target nucleic acid. Accordingly, in one aspect, provided herein is a composition comprising an effector protein and a guide nucleic acid for the treatment of one or more disorders associated with irregular expression of the functional human protein, for instance, monogenic diseases.

Certain Embodiments

[6] Provided herein are compositions comprising: an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**, wherein the guide nucleic acid comprises a spacer sequence that hybridizes to a target sequence in a target nucleic acid, wherein the target sequence comprises a nucleotide sequence within a human safe harbor locus, and wherein the donor nucleic acid encodes a transgene that comprises a functional human protein that is expressed upon incorporation into the human safe harbor locus.

[7] Also provided herein the compositions comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the effector protein comprises an amino acid sequence with at least 95% sequence identity to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the effector protein comprises one or more amino acid substitutions relative to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the one or more amino acid substitutions independently comprise one or more conservative substitutions, one or more non-conservative substitutions, or combinations thereof. In some embodiments, the one or more amino acid substitutions comprise one or more substitutions with a positively charged amino acid residues. In some embodiments, the positively charged amino acid residue is independently selected from Lys (K), Arg (R), and His (H).

[8] Also provided herein are compositions comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the effector protein recognizes a protospacer adjacent motif (PAM) sequence comprising any one of the nucleotide sequences recited in **TABLE 3**. In some embodiments, the effector protein recognizes a target sequence that is in proximity to or adjacent to a protospacer adjacent motif (PAM) sequence comprising any one of the nucleotide sequences recited in **TABLE 3**.

[9] Also provided herein are compositions comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the effector protein is fused to a nuclear localization signal.

[10] Also provided herein are compositions comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the composition further comprises a fusion partner protein linked to the effector protein. In some embodiments, the fusion partner protein is directly fused to the N terminus or C terminus of the effector protein via an amide bond. In some embodiments, the nuclear localization signal comprises an amino acid sequence comprising any one of the amino acid sequences recited in **TABLE 2**.

[11] Provided herein are compositions comprising: an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the guide nucleic acid comprises a spacer sequence that hybridizes to a target sequence in a target nucleic acid, wherein the target sequence comprises a nucleotide sequence within a human safe harbor locus, and wherein the human safe harbor locus is an intron. In certain embodiments, the human safe harbor comprises a genomic safe harbor sequence that upon incorporation of the transgene results in expression of the functional human protein without perturbing any endogenous genes.

[12] Provided herein are compositions comprising: an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the guide nucleic acid comprises a spacer sequence that hybridizes to a target sequence in a target nucleic acid, wherein the target sequence comprises a nucleotide sequence within a human safe harbor locus, and wherein the human safe harbor locus is a gene or portion thereof that is expressed in the liver.

[13] Provided herein are compositions comprising: an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the guide nucleic acid comprises a spacer sequence that hybridizes to a target sequence in a target nucleic acid, wherein the target sequence comprises a nucleotide sequence within a human safe harbor locus, and wherein the human safe harbor locus comprises any one of the nucleotide sequences recited **TABLE 10**. Provided herein are compositions comprising: an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the guide nucleic acid comprises a spacer sequence that hybridizes to a target sequence in a target nucleic acid, wherein the target sequence comprises a nucleotide sequence within a human safe harbor locus, and wherein the human safe harbor locus comprises at least one sequence in AAVS1, CCR5, hRosa26 or human serum albumin (HSA).

[14] Provided herein are compositions comprising: an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the guide nucleic acid comprises a spacer sequence that hybridizes to a target sequence in a target nucleic acid, wherein the target sequence comprises a nucleotide sequence within a human safe harbor locus, and wherein the human safe harbor locus is located in human chromosome 1, human chromosome 3, human chromosome 4, human chromosome 6, human chromosome 10, human chromosome 11, human chromosome 12, human chromosome 14, human chromosome 17, human chromosome 18, or human chromosome 19. Provided herein are compositions comprising: an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the guide nucleic acid comprises a spacer sequence that hybridizes to a target sequence in a target nucleic acid, wherein the target sequence

comprises a nucleotide sequence within a human safe harbor locus, and wherein the human safe harbor locus is located in human chromosome 2 or human chromosome 4.

[15] Also provided herein are compositions comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the guide nucleic acid comprises a nucleotide sequence that is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 4**, **TABLE 5**, **TABLE 6**, **TABLE 7**, **TABLE 9** and any combination thereof. In some embodiments, the guide nucleic acid comprises a spacer sequence that is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 4** and **TABLE 5**.

[16] Also provided herein are compositions comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid. In some embodiments, the guide nucleic acid comprises a repeat sequence that is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 6**. Also provided herein are compositions comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the guide nucleic acid comprises a repeat sequence that is at least 90% identical to **SEQ ID NO: 515** or **516**.

[17] Also provided herein are compositions comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the guide nucleic acid does not comprise a tracrRNA.

[18] Also provided herein are compositions comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the guide nucleic acid is a single guide RNA (sgRNA).

[19] Also provided herein are compositions comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the guide nucleic acid is a crRNA.

[20] Also provided herein are compositions comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the guide nucleic acid comprises a handle RNA sequence that is at least 90% identical to any one of **SEQ ID NO: 259-264**.

[21] Also provided herein are compositions comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the target sequence is a human gene.

[22] Also provided herein are compositions comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the target sequence comprises a nucleotide sequence that is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 10**.

[23] Also provided herein are compositions comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the composition results in cleavage of both strands of the human safe harbor locus prior to incorporation of the donor nucleic acid.

[24] Also provided herein are compositions wherein the functional human protein is select from any one of A1AT, CFTR, DMD, FXN, F8, F9, GAA, SOD1, C9, HTT, MECP2, SMN1, TARDBP, FUS, RHO, and USH2A or a functional variant or fragment thereof.

[25] Also provided herein are compositions comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the guide nucleic acid comprises a spacer sequence that hybridizes to a target sequence in a target nucleic acid, and wherein the target sequence comprises a nucleotide sequence within **SEQ ID NO: 266**.

[26] Also provided herein are compositions comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the guide nucleic acid comprises a spacer sequence, wherein the spacer sequence is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 4**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6, 13, 47,** and **228-231**. In some embodiments, the guide nucleic acid comprises a repeat sequence of **SEQ ID NO: 252** or **1848**, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6** and **228-230**, and wherein the target sequence is adjacent to a PAM of **SEQ ID NO: 234-237**. In some embodiments, the guide nucleic acid comprises a repeat sequence of **SEQ ID NO: 253**, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 13**, and wherein the target sequence is adjacent to any one of a PAM of **SEQ ID NO: 238, 240-247**. In some embodiments, the guide nucleic acid comprises a handle sequence of **SEQ ID NO: 259-261**, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 13**, and wherein the target sequence is adjacent to a PAM of **SEQ ID NO: 238, 240-247**. In some embodiments, the guide nucleic acid comprises a repeat sequence of **SEQ ID NO: 254**, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 47**, and wherein the target sequence is adjacent to a PAM of **SEQ ID NO: 241-244**. In some embodiments, the guide nucleic acid comprises a repeat sequence of **SEQ ID NO: 255** or **1789**, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 231**, and wherein the target sequence is adjacent to a PAM of any one of **SEQ ID NO: 245-248, 1829,** and **1834-1844**. In some embodiments, the guide nucleic acid comprises any one of handle sequences of **SEQ ID NO: 262-264**, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 231**, and wherein the target sequence is adjacent to a PAM of any one of **SEQ ID NO: 245-248, 1829,** and **1834-1844**. In some embodiments, the guide nucleic acid comprises a spacer sequence that is at least 90% identical to

any one of **SEQ ID NO: 348, 558, 401, 409, 473** and **486**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6,** and **228-230**. In some embodiments, the guide nucleic acid comprises a spacer sequence that is at least 90% identical to any one of **SEQ ID NO: 592, 600** and **602**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 231**.

[27] Also provided herein are compositions comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the guide nucleic acid comprises a nucleotide sequence that is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 9**. In some embodiments, the guide nucleic acid comprises a nucleotide sequence that is at least 90% identical to any one of **SEQ ID NO: 652, 685, 705, 713, 777** and **790**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6,** and **228-230**. In some embodiments, the guide nucleic acid comprises a nucleotide sequence that is at least 90% identical to any one of **SEQ ID NO: 856, 864** and **866**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical **SEQ ID NO: 231**.

[28] Also provided herein are compositions comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the guide nucleic acid comprises a spacer sequence that hybridizes to a target sequence in a target nucleic acid, and wherein the target sequence comprises a nucleotide sequence within **SEQ ID NO: 1795**.

[29] Also provided herein are compositions comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the guide nucleic acid comprises a spacer sequence, wherein the spacer sequence is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 5**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6, 13, 47,** and **228-231**.

[30] Also provided herein the compositions comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**, wherein the guide nucleic acid is a crRNA, wherein the crRNA comprises a spacer sequence and a repeat sequence, wherein the spacer sequence hybridizes to a target sequence in a target nucleic acid, wherein the target sequence comprises a nucleotide sequence within a human safe harbor locus, wherein the target sequence comprises any one of the nucleotide sequences recited in **TABLE 10**, and wherein the donor nucleic acid encodes a functional human protein that upon expression after introduction in a human subject expression replaces the function of a corresponding nonworking or missing gene in the

human subject. In some embodiments, the repeat sequence is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 6.** In some embodiments, the repeat sequence is identical to any one of **SEQ ID NO: 252-258, 1789,** and **1848**. Also provided herein the compositions comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**, wherein the guide nucleic acid is a crRNA, wherein the crRNA comprises a spacer sequence and a repeat sequence, wherein the spacer sequence hybridizes to a target sequence in a target nucleic acid, wherein the repeat sequence is identical to any one of **SEQ ID NO: 252-258, 1789,** and **1848**, wherein the target sequence comprises a nucleotide sequence within a human safe harbor locus, and wherein the donor nucleic acid encodes a functional human protein that upon expression after introduction in a human subject expression replaces the function of a corresponding nonworking or missing gene in the human subject. In some embodiments, the effector protein comprises one or more amino acid substitutions relative to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the one or more amino acid substitutions independently comprise one or more conservative substitutions, one or more non-conservative substitutions, or combinations thereof. In some embodiments, the one or more amino acid substitutions comprise one or more substitutions with a positively charged amino acid residues. In some embodiments, the positively charged amino acid residue is independently selected from Lys (K), Arg (R), and His (H). In some embodiments, the functional human protein is select from any one of A1AT, CFTR, DMD, FXN, F8, F9, GAA SOD1, C9, HTT, MECP2, SMN1, TARDBP, FUS, RHO, and USH2A or a functional variant or fragment thereof. In some embodiments, the composition is formulated for administration to a human subject suffering from a disorder characterized by irregular expression of the functional human protein absent the administration. In some embodiments, the target sequence comprises any one of the nucleotide sequences recited in **TABLE 10**. In some embodiments, the spacer sequence is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 4** and **TABLE 5**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6, 47,** and **228-230**. In some embodiments, the target sequence comprises a nucleotide sequence within **SEQ ID NO: 266**. In some embodiments, the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6, 47,** and **228-230.** In some embodiments, the spacer sequence is at least 90% identical to any one of **SEQ ID NO: 301-510** and **536-575**. In some embodiments, the spacer sequence is at least 90% identical to any one of **SEQ ID NO: 348, 401, 409, 473, 486,** and **558**, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6,** and **228-230**. In some embodiments, the spacer sequence is at least 90% identical to any one of **SEQ ID NO: 301-510** and **536-575**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6, 47,** and **228-230**. In some embodiments, the crRNA comprises a nucleotide sequence that is at least at least 90%

identical to any one of **SEQ ID NO: 605-839**. In some embodiments, the crRNA comprises a nucleotide sequence that is at least 90% identical to any one of **SEQ ID NO: 652, 685, 705, 713, 777** and **790**, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6,** and **228-230**. In some embodiments, the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6,** and **228-230**, wherein the effector protein comprising one or more amino acid substitutions independently selected from E109R, H208R, K184R, K38R, L182R, Q183R, S108R, S198R, T114R or a combination thereof.

[31] Also provided herein are composition comprising an effector protein, or a nucleic acid encoding the effector protein; a guide nucleic acid, or a nucleic acid encoding the guide nucleic acid; and a donor nucleic acid, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 6**, wherein the guide nucleic acid is a single guide RNA (sgRNA), wherein the sgRNA comprises a spacer sequence and a handle sequence, wherein the spacer sequence hybridizes to a target sequence in a target nucleic acid, wherein the handle sequence comprises an intermediary sequence, wherein the target sequence comprises a nucleotide sequence within a human safe harbor locus, and wherein the donor nucleic acid encodes a functional human protein that upon expression after introduction in a human subject expression replaces the function of a corresponding nonworking or missing gene in the human subject. In some embodiments, the handle sequence further comprises one or more of a repeat sequence and a linker, wherein the repeat sequence is identical to **SEQ ID NO: 253** or **255**.

[32] Also provided herein are composition comprising an effector protein, or a nucleic acid encoding the effector protein; a guide nucleic acid, or a nucleic acid encoding the guide nucleic acid; and a donor nucleic acid, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 6**, wherein the guide nucleic acid is a single guide RNA (sgRNA), wherein the sgRNA comprises a spacer sequence and a handle sequence, wherein the spacer sequence hybridizes to a target sequence in a target nucleic acid, wherein the handle sequence comprises one or more of a tracrRNA sequence or a portion thereof, a repeat sequence, and a linker, wherein the repeat sequence is identical to **SEQ ID NO: 253** or **255**, wherein the target sequence comprises a nucleotide sequence within a human safe harbor locus, and wherein the donor nucleic acid encodes a functional human protein that upon expression after introduction in a human subject expression replaces the function of a corresponding nonworking or missing gene in the human subject.

[33] Also provided herein are composition comprising an effector protein, or a nucleic acid encoding the effector protein; a guide nucleic acid, or a nucleic acid encoding the guide nucleic acid; and a donor nucleic acid, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**, wherein the guide nucleic acid is a single guide RNA (sgRNA), wherein the sgRNA comprises a spacer sequence and a handle sequence, wherein the spacer sequence hybridizes to a target sequence in a target nucleic acid, wherein the handle sequence comprises an intermediary sequence, wherein the target sequence

comprises a nucleotide sequence within a human safe harbor locus, wherein the target sequence comprises any one of the nucleotide sequences recited in **TABLE 10,** and wherein the donor nucleic acid encodes a functional human protein that upon expression after introduction in a human subject expression replaces the function of a corresponding nonworking or missing gene in the human subject. In some embodiments, the effector protein comprises one or more amino acid substitutions relative to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the one or more amino acid substitutions independently comprise one or more conservative substitutions, one or more non-conservative substitutions, or combinations thereof. In some embodiments, the one or more amino acid substitutions comprise one or more substitutions with a positively charged amino acid residues. In some embodiments, the positively charged amino acid residue is independently selected from Lys (K), Arg (R), and His (H). In some embodiments, the spacer sequence is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 4** and **TABLE 5**. In some embodiments, the repeat sequence is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 6.** In some embodiments, the repeat sequence is identical to any one of **SEQ ID NO: 252-258, 1789,** and **1848**. In some embodiments, the handle is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 7.** In some embodiments, the handle is identical in a range of from 90% to 100% to any one of **SEQ ID NO: 259-264**. In some embodiments, the linker sequence comprises a nucleotide sequence that is identical to the nucleotide sequence recited in **TABLE 8.** In some embodiments, the linker sequence comprises a nucleotide sequence that is identical to **SEQ ID NO: 265.** In some embodiments, the target sequence comprises a nucleotide sequence within **SEQ ID NO: 266**. In some embodiments, the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 13** and **231.** In some embodiments, the spacer sequence is at least 90% identical to any one of **SEQ ID NO: 536-604**. In some embodiments, the spacer sequence is at least 90% identical to any one of **SEQ ID NO: 592, 600** and **602**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 231**. In some embodiments, the sgRNA comprises a nucleotide sequence that is at least 90% identical to any one of **SEQ ID NO: 840-908**. In some embodiments, the sgRNA comprises a nucleotide sequence that is at least 90% identical to any one of **SEQ ID NO: 856, 864** and **866**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 231**. In some embodiments, the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 231**, wherein the effector protein comprises one or more amino acid substitutions independently selected from K58W, I80K, N193K, S209F, A218K, E225K, N286K, M295W, M298L, A306K, Y315M or a combination thereof. In some embodiments, the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 231**, wherein the effector protein comprises one or more amino acid substitutions independently selected from I80R, T84R, K105R, G210R, C202R, A218R, D220R, E225R, C246R, Q360R or a combination thereof.

[34] Also provided herein the compositions comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the handle sequence is identical in a range of from 90% to 100% to any one of **SEQ ID NO: 259-264**. In some embodiments, the target sequence comprises any one of the nucleotide sequences recited in **TABLE 10**. In some embodiments, the spacer sequence is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 4** and **TABLE 5**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 13** and **231**. In some embodiments, the target sequence comprises a nucleotide sequence within **SEQ ID NO: 266**. In some embodiments, the spacer sequence is at least 90% identical to any one of **SEQ ID NO: 536-604**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 13** and **231**. In some embodiments, the sgRNA comprises a nucleotide sequence that is at least 90% identical to any one of **SEQ ID NO: 840-908**.

[35] Also provided herein are compositions comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the guide nucleic acid is identical in a range of from 90% to 100% to any one of **SEQ ID NO: 815-839**. In some embodiments, the crRNA is at least 90% identical to any one of **SEQ ID NO: 551-575**.

[36] Also provided herein the compositions comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the composition comprises an expression vector, wherein the expression vector comprises at least one of the nucleic acid encoding the effector protein; the nucleic acid encoding the guide nucleic acid; and the donor nucleic acid.

[37] Also provided herein the compositions comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the composition comprises a viral vector.

[38] Also provided herein the compositions comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the composition comprises an adeno associated viral (AAV) vector.

[39] Also provided herein the compositions comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the nucleic acid encoding the effector protein is a messenger RNA.

[40] Also provided the compositions described herein, wherein the composition comprising a lipid or a lipid nanoparticle.

[41] Also provided herein are pharmaceutical compositions, wherein the pharmaceutical compositions comprise the compositions described herein and a pharmaceutically acceptable excipient.

[42] Also provided herein are systems comprising components for introduction of a donor nucleic acid encoding a functional human protein into a human safe harbor locus, wherein the components

comprise an effector protein, or a nucleic acid encoding the effector protein, wherein the effector protein comprises an amino acid sequence having at least 90% sequence identity to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**; a guide nucleic acid, or a nucleic acid encoding the guide nucleic acid, wherein the guide nucleic acid comprises a spacer sequence that hybridizes to a target sequence in a target nucleic acid, and wherein the target sequence comprises a nucleotide sequence within the human safe harbor locus; and the donor nucleic acid. In some embodiments, the effector protein comprises one or more amino acid substitutions relative to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the one or more amino acid substitutions independently comprise one or more conservative substitutions, one or more non-conservative substitutions, or combinations thereof. In some embodiments, the one or more amino acid substitutions comprise one or more substitutions with a positively charged amino acid residues. In some embodiments, the positively charged amino acid residue is independently selected from Lys (K), Arg (R), and His (H).

[43] Also provided herein are systems wherein the target sequence is in proximity to or adjacent to a protospacer adjacent motif (PAM) sequence comprising any one of the nucleotide sequences recited in **TABLE 3**, and wherein the effector protein and guide form a complex that modifies the human safe harbor locus upon hybridization of the guide nucleic acid to the target sequence and recognition of the PAM sequence by the effector protein. In some embodiments, the target sequence comprises any one of the nucleotide sequences recited in **TABLE 10**. In some embodiments, the spacer sequence is at least 90% identical to any one of nucleotide sequences recited in **TABLE 4** and **TABLE 5**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6, 13, 47,** and **228-231**. In some embodiments, the target sequence comprises a nucleotide sequence within **SEQ ID NO: 266**. In some embodiments, the spacer sequence is at least 90% identical to any one of nucleotide sequences recited in **TABLE 4**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6, 13, 47,** and **228-231**. In some embodiments, the guide nucleic acid comprises a nucleotide sequence that is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 9**.

[44] Also provided herein are methods of expressing a functional human protein in a cell, the method comprising contacting the cell with the compositions or systems described herein.

[45] Also provided herein are methods of inserting a donor nucleic acid within a safe harbor loci using CasPhi.12 L26R or variants thereof. In some embodiments, the donor nucleic acid comprises a nucleotide sequence that encodes a functional human protein into a safe harbor locus.

[46] Also provided herein are methods of treating a disease associated with a mutation or aberrant expression of a functional human protein in a subject in need thereof, wherein the method comprising administering to the subject the composition described herein. In some embodiments, the functional human protein can be any one of proteins described in **TABLE 11**.

[47] Also provided herein are methods of treating a disease associated with a mutation or aberrant expression of a functional human protein in a subject in need thereof, wherein the method comprising administering to the subject the composition described herein. In some embodiments, the functional human protein can be any one of proteins described in **TABLE 11**.

[48] Also provided herein are methods of treating a disease associated with a mutation or aberrant expression of a human protein in a subject in need thereof, wherein the subject has a genetic disorder. In some embodiments, the genetic disorder is a monogenic disorder. In some embodiments, the disease or disorder is one or more of glycogen storage disorder, cystic fibrosis, muscular dystrophy, Freidreich's ataxia, amylotrophic lateral scleorsis, hemophilia, Huntington's disease, retinal dystrophy, Rett syndrome, and sickle cell disease.

[49] Also provided herein are methods of treating a disease associated with a mutation or aberrant expression of a human protein. In some embodiments, the subject has a reduced activity of the human protein prior to the administering. In some embodiments, the subject has no activity of the human protein prior to the administering. In some embodiments, administering increases the amount of the functional human protein in a biological sample of the subject relative to the amount or concentration of functional protein in the biological sample before administering. In some embodiments, the amount or concentration of functional protein in the biological sample is increased by at least 10%, at least 20%, at least 30%, at least 40%, or at least 50%. In some embodiments, the biological sample is selected from blood, serum, plasma, urine, saliva, and cerebrospinal fluid.

[50] Also provided herein are methods of treating a disease associated with a mutation or aberrant expression of a human protein in a subject in need thereof, wherein the subject has one or more genetic mutations.

[51] Also provided herein are methods of treating a disease associated with a mutation or aberrant expression of a human protein in a subject in need thereof, wherein the mutation comprises a point mutation, a single nucleotide polymorphism (SNP), a chromosomal mutation, a copy number mutation, or any combination thereof.

[52] Also provided methods of treating described herein, wherein the mutation is associated with one or more of protein expression, protein activity, and protein structural stability.

[53] Also provided herein are methods of treating a disease associated with a mutation or aberrant expression of a protein in a subject in need thereof, wherein the mutation results in reduced expression of the protein, reduced enzymic activity, reduced half-life of the protein or a combination thereof, in the cell relative to a cell without the mutation, before the treatment.

[54] Also provided herein are methods of treating a disease associated with a mutation or aberrant expression of a protein in a subject in need thereof, wherein the method is performed in cell.

[55] Also provided herein are methods of expressing a protein or methods of treating described herein, wherein the method is performed in vivo.

[56] Also provided herein is a cell comprising compositions described herein.

[57] Also provided herein is a cell that comprises the target nucleic acid modified by compositions described herein.

[58] Also provided cells described herein, wherein the cell is a hepatocyte.

[59] Also provided cells described herein, wherein the cell is a mammalian cell.

[60] Also provided cells described herein, wherein the cell is a human cell.

[61] Also provided cells described herein, wherein the cell is a: stem cell, progenitor cell, induced pluripotent stem cell (iPSC) or a cell derived from an iPSC.

[62] Also provided herein are populations of cells that comprises at least one cell described herein.

[63] Also provided herein are systems for introduction of a donor nucleic acid into a safe harbor. In some embodiments, the systems comprise one or more components, wherein the one or more components individually comprises one or more of the following: (i) an effector protein, or a nucleic acid encoding the effector protein, wherein the effector protein comprises an amino acid sequence having at least 90% sequence identity to any one of amino acid sequences recited in **TABLE 1A** and **TABLE 1B**; (ii) a guide nucleic acid, or a nucleic acid encoding the guide nucleic acid, wherein the guide nucleic acid comprises a spacer sequence that hybridizes to a target sequence in a target nucleic acid, and wherein the target sequence comprises a nucleotide sequence within the safe harbor; and (iii) the donor nucleic acid encoding a transgene that comprises a functional human protein that is expressed upon incorporation into the human safe harbor locus, wherein the safe harbor comprises at least 90% sequence identity to any one of nucleotide sequences recited in **TABLE 10**. In some embodiments, the effector protein comprises one or more amino acid substitutions relative to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the one or more amino acid substitutions independently comprise one or more conservative substitutions, one or more non-conservative substitutions, or combinations thereof. In some embodiments, the one or more amino acid substitutions comprise one or more substitutions with a positively charged amino acid residues. In some embodiments, the positively charged amino acid residue is independently selected from Lys (K), Arg (R), and His (H). In some embodiments, the effector protein comprises an amino acid sequence with at least 95% sequence identity to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the target sequence is in proximity to or adjacent to a protospacer adjacent motif (PAM) sequence comprising any one of the nucleotide sequences recited in **TABLE 3**. In some embodiments, the effector protein comprises a nuclear localization signal. In some embodiments, the system further comprises a fusion partner protein linked to the effector protein. In some embodiments, the fusion partner protein is directly fused to the N terminus or C terminus of the effector protein via an amide bond. In some embodiments, the human safe harbor locus is located in human chromosome 1, human chromosome 3, human chromosome 4, human chromosome 6, human chromosome 10, human chromosome 11, human chromosome 12, human chromosome 14, human chromosome 17, human chromosome 18, or human chromosome 19. In some embodiments, the guide nucleic acid is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 4**, **TABLE 5**, **TABLE 6**,

TABLE 7, **TABLE 9** and any combination thereof. In some embodiments, the guide nucleic acid comprises a spacer sequence that is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 4** and **TABLE 5**. In some embodiments, the guide nucleic acid does not comprise a tracrRNA. In some embodiments, the system is capable of cleaving both strands of the human safe harbor locus prior to incorporation of the donor nucleic acid. In some embodiments, the functional human protein is select from any one of CFTR, DMD, GAA, A1AT, FXN, F8, F9, SOD1, C9, HTT, MECP2, SMN1, TARDBP, FUS, RHO, and USH2A or a functional variant or fragment thereof. In some embodiments, the target sequence comprises a nucleotide sequence within **SEQ ID NO: 266**. In some embodiments, the spacer sequence is at least 90% identical to any one of nucleotide sequences recited in **TABLE 4**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6, 13, 47**, and **228-231**. In some embodiments, the guide nucleic acid comprises a spacer sequence that is at least 90% identical to any one of **SEQ ID NO: 348, 558, 401, 409, 473** and **486**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6** and **228-230**. In some embodiments, the guide nucleic acid comprises a spacer sequence that is at least 90% identical to any one of **SEQ ID NO: 592, 600** and **602**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 231**. In some embodiments, the guide nucleic acid comprises a nucleotide sequence that is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 9**. In some embodiments, the guide nucleic acid comprises a nucleotide sequence that is at least 90% identical to any one of **SEQ ID NO: 652, 685, 705, 713, 777** and **790**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6,** and **228-230**. In some embodiments, the guide nucleic acid comprises a nucleotide sequence that is at least 90% identical to any one of **SEQ ID NO: 856, 864** and **866**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 231**. In some embodiments, the target sequence comprises a nucleotide sequence within **SEQ ID NO: 1795**. In some embodiments, the spacer sequence is at least 90% identical to any one of nucleotide sequences recited in **TABLE 5**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6, 13, 47,** and **228-231**. In some embodiments, the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6, 228-230**, wherein the effector protein comprises one or more amino acid substitutions independently selected from E109R, H208R, K184R, K38R, L182R, Q183R, S108R, S198R, T114R, or a combination thereof. In some embodiments, the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 231**, wherein the effector protein comprises one or more amino acid substitutions independently selected from K58W, I80K, N193K, S209F, A218K, E225K, N286K, M295W, M298L, A306K, Y315M, or a combination thereof. In some embodiments, the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 231**, wherein the effector protein

comprises one or more amino acid substitutions independently selected from I80R, T84R, K105R, G210R, C202R, A218R, D220R, E225R, C246R, Q360R, or a combination thereof.

[64] Also provided herein the systems comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the system comprises a viral vector.

[65] Also provided herein the systems comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the system comprises an adeno associated viral (AAV) vector.

[66] Also provided herein the systems comprising an effector protein or a nucleic acid encoding the effector protein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid, and a donor nucleic acid, wherein the nucleic acid encoding the effector protein is a messenger RNA.

[67] Also provided the systems described herein, wherein the system comprising a lipid or a lipid nanoparticle.

[68] Also provided the systems described herein, wherein any two components of the system are provided in different solutions or containers. In some embodiments, at least two components of the system are administered separately or simultaneously.

[69] Also provided the composition, the pharmaceutical composition, the system, the method, or the cells described herein, wherein the target nucleic acid is within the human albumin gene.

[70] Also provided the composition, the system, the method or the cells described herein, wherein the target nucleic acid is at least partially within a targeted intron within the human albumin gene.

[71] Also provided the composition, the system, the method or the cells described herein, wherein at least a portion of the target nucleic acid that the guide nucleic acid binds to comprises about 30 nucleotides to about 150 nucleotides adjacent to: the start of the intron 1, the end of the intron 1, or both.

INCORPORATION BY REFERENCE

[72] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[73] **FIGs. 1A-1F** shows indel activity of the effector protein CasPhi.12 L26R **(SEQ ID NO: 228)** for gRNA within intron 1 of human albumin, as described in Example 1;

[74] **FIG. 2** shows indel activity of the effector proteins, CasPhi.12 **(SEQ ID NO: 6)** and CasPhi.12 L26R **(SEQ ID NO: 228)**, for gRNA within intron 1 of human albumin, as described in Example 1;

[75] **FIG. 3** shows indel activity of the effector protein CasPhi.32 **(SEQ ID NO: 47)** for gRNA within intron 1 of human albumin, as described in Example 2;

[76] **FIG. 4** shows indel activity of the effector protein CasM.265466 (**SEQ ID NO: 231**) for gRNA within intron 1 of human albumin, as described in Example 3; and

[77] **FIG. 5** shows indel activity of the effector protein CasPhi.12 L26R (**SEQ ID NO: 228**) for gRNA within intron 1 of human albumin, as described in Example 6.

[78] **FIG. 6A-6B** shows the nuclease activity of CasM.265466 with flexible PAM sequences, in accordance with an embodiment of the present disclosure.

[79] **FIG. 7A-7B** illustrates the effects of an arginine substitution on CasM.265466 nuclease activity for a target nucleic acid, in accordance with an embodiment of the present disclosure.

[80] **FIG. 8** illustrates the dose titration curves of CasM.265466 arginine mutants, in accordance with an embodiment of the present disclosure.

[81] **FIG. 9** illustrates %indel generated by D220R effector protein variant of CasM.265466 relative to corresponding wildtype CasM.265466 and Cas9 effector proteins.

[82] **FIG. 10** shows % indel generated at various dose (ng) for D220R and A306K variants relative to CasM.265466 and Cas9 effector proteins.

[83] **FIG. 11** shows the activity of the engineered variants relative to that of wildtype CasM.265466 (**SEQ ID NO: 231**) as fold change.

[84] **FIG. 12** shows that CasPhi.12 variant enzymes can bind two genome loci of mammalian cells and edit the genome at the locus with varying efficacy normalized to the wild-type (**SEQ ID NO: 6**). The x and y-axis of the plot corresponds to various targeted loci. The identifier next to each plotted data point denotes the amino acid residue alteration and position in reference to **SEQ ID NO: 6**.

[85] **FIGs. 13A-13B** show indel activity of CasPhi.12 variant enzymes. **FIG. 13A** shows indel activity of CasPhi.12 variant enzymes. **FIG. 13B** shows indel activity of CasPhi.12 variant enzymes normalized to WT (**SEQ ID NO: 6**). The identifier under to each pbar denotes the amino acid residue alteration and position in reference to **SEQ ID NO: 6**.

[86] **FIG. 14** shows gel electrophoresis analysis of *cis* cleavage activity by CasPhi.12 and variants thereof.

[87] **FIG. 15** shows gel electrophoresis analysis of *cis* cleavage activity by CasM.265466 and variants thereof.

[88] **FIG. 16A** shows indel activity of the effector protein system (*e.g.*, CasPhi.12 L26R, CasM.265466) within or adjacent to intron 1 of human albumin gene as related to different concentrations of RNA and MOI and as compared to positive control (*e.g.*, SpyCas9). **FIG. 16B** shows relative light units (RLU) as a measure of integration activity of the effector protein system (*e.g.*, CasPhi.12 L26R, CasM.265466) within or adjacent to intron 1 of human albumin gene as related to different concentrations of RNA and MOI and as compared to positive control (*e.g.*, SpyCas9).

[89] **FIG. 17** shows % integration products as a measure of integration activity of the effector protein system (*e.g.*, CasPhi.12 L26R, CasM.265466) within or adjacent to intron 1 of human albumin gene as

compared to positive control (*e.g.*, SpyCas9) as measured via reverse transcription droplet digital PCR (RT-ddPCR).

DETAILED DESCRIPTION OF THE INVENTION

[90] It is to be understood that both the foregoing general description and the following detailed description are exemplary, and explanatory only, and are not restrictive of the disclosure.

[91] The section headings used herein are for organizational purposes only and are not to be construed as limiting the subject matter described.

[92] All documents, or portions of documents, cited in this application, including, but not limited to, patents, patent applications, articles, books, and treatises, are hereby expressly incorporated by reference in their entirety for any purpose.

Definitions

[93] Unless otherwise indicated, all technical terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Unless otherwise indicated or obvious from context, the following terms have the following meanings:

[94] The terms, "% identical," "% identity," and "percent identity," or grammatical equivalents thereof, refer to the extent to which two sequences (nucleotide or amino acid) have the same residue at the same positions in an alignment. For example, "an amino acid sequence is X% identical to SEQ ID NO: Y" can refer to % identity of the amino acid sequence to SEQ ID NO: Y and is elaborated as X% of residues in the amino acid sequence are identical to the residues of sequence disclosed in SEQ ID NO: Y. Generally, computer programs can be employed for such calculations. Illustrative programs that compare and align pairs of sequences, include ALIGN (Myers and Miller, *Comput Appl Biosci.* 1988 Mar;4(1):11-7), FASTA (Pearson and Lipman, *Proc Natl Acad Sci U S A.* 1988 Apr;85(8):2444-8; Pearson, *Methods Enzymol.* 1990;183:63-98) and gapped BLAST (Altschul *et al., Nucleic Acids Res.* 1997 Sep 1;25(17):3389-40), BLASTP, BLASTN, or GCG (Devereux *et al., Nucleic Acids Res.* 1984 Jan 11;12(1 Pt 1):387-95).

[95] The terms, "amplification" and "amplifying," or grammatical equivalents thereof, as used herein, refer to a process by which a nucleic acid molecule is enzymatically copied to generate a plurality of nucleic acid molecules containing the same sequence as the original nucleic acid molecule or a distinguishable portion thereof.

[96] The term, "base editing enzyme," as used herein, refers to a protein, polypeptide or fragment thereof that is capable of catalyzing the chemical modification of a nucleobase of a deoxyribonucleotide or a ribonucleotide. Such a base editing enzyme, for example, is capable of catalyzing a reaction that modifies a nucleobase that is present in a nucleic acid molecule, such as DNA or RNA (single stranded or double stranded). Non-limiting examples of the type of modification that a base editing enzyme is capable of catalyzing includes converting an existing nucleobase to a different nucleobase, such as converting a cytosine to a guanine or thymine or converting an adenine to a guanine, hydrolytic

deamination of an adenine or adenosine, or methylation of cytosine (*e.g.*, CpG, CpA, CpT or CpC). A base editing enzyme itself may or may not bind to the nucleic acid molecule containing the nucleobase.

[97] The term, "base editor," as used herein, refers to a fusion protein comprising a base editing enzyme fused to an effector protein. The base editor is functional when the effector protein is coupled to a guide nucleic acid. The guide nucleic acid imparts sequence specific activity to the base editor. By way of non-limiting example, the effector protein may comprise a catalytically inactive effector protein. Also, by way of non-limiting example, the base editing enzyme may comprise deaminase activity. Additional base editors are described herein.

[98] The term, "catalytically inactive effector protein," as used herein, refers to an effector protein that is modified relative to a naturally-occurring effector protein to have a reduced or eliminated catalytic activity relative to that of the naturally-occurring effector protein, but retains its ability to interact with a guide nucleic acid. The catalytic activity that is reduced or eliminated is often a nuclease activity. The naturally-occurring effector protein may be a wildtype protein. In some embodiments, the catalytically inactive effector protein is referred to as a catalytically inactive variant of an effector protein, *e.g.*, a Cas effector protein.

[99] The term, "*cis* cleavage," as used herein, refers to cleavage (hydrolysis of a phosphodiester bond) of a target nucleic acid by an effector protein complexed with a guide nucleic acid *(e.g.,* an RNP complex), wherein at least a portion of the guide nucleic acid is hybridized to the target nucleic acid. Cleavage may occur within or directly adjacent to the region of the target nucleic acid that is hybridized to the guide nucleic acid.

[100] The terms, "complementary" and "complementarity," as used herein, with reference to a nucleic acid molecule or nucleotide sequence, refer to the characteristic of a polynucleotide having nucleotides that base pair with their Watson-Crick counterparts (C with G; or A with T) in a reference nucleic acid. For example, when every nucleotide in a polynucleotide forms a base pair with a reference nucleic acid, that polynucleotide is said to be 100% complementary to the reference nucleic acid. In a double stranded DNA or RNA sequence, the upper (sense) strand sequence is in general, understood as going in the direction from its 5′- to 3′-end, and the *complementary* sequence is thus understood as the sequence of the lower (antisense) strand in the same direction as the upper strand. Following the same logic, the *reverse* sequence is understood as the sequence of the upper strand in the direction from its 3′- to its 5′-end, while the '*reverse complement*' sequence or the '*reverse complementary*' sequence is understood as the sequence of the lower strand in the direction of its 5′- to its 3′-end. Each nucleotide in a double stranded DNA or RNA molecule that is paired with its Watson-Crick counterpart called its *complementary* nucleotide.

[101] The terms, "cleave," "cleaving," and "cleavage," as used herein, with reference to a nucleic acid molecule or nuclease activity of an effector protein, refer to the hydrolysis of a phosphodiester bond of a nucleic acid molecule that results in breakage of that bond. The result of this breakage can be a nick (hydrolysis of a single phosphodiester bond on one side of a double-stranded molecule), single

strand break (hydrolysis of a single phosphodiester bond on a single-stranded molecule) or double strand break (hydrolysis of two phosphodiester bonds on both sides of a double-stranded molecule) depending upon whether the nucleic acid molecule is single-stranded (*e.g.*, ssDNA or ssRNA) or double-stranded (*e.g.*, dsDNA) and the type of nuclease activity being catalyzed by the effector protein.

[102] The term, "clustered regularly interspaced short palindromic repeats (CRISPR)," as used herein, refers to a segment of DNA found in the genomes of certain prokaryotic organisms, including some bacteria and archaea, that includes repeated short sequences of nucleotides interspersed at regular intervals between unique sequences of nucleotides derived from the DNA of a pathogen (*e.g.*, virus) that had previously infected the organism and that functions to protect the organism against future infections by the same pathogen.

[103] The terms, "CRISPR RNA" and "crRNA," as used herein, refer to a type of guide nucleic acid, wherein the nucleic acid is RNA comprising a first sequence, often referred to herein as a spacer sequence, that hybridizes to a target sequence of a target nucleic acid, and a second sequence that is capable of connecting a crRNA to an effector protein by either a) hybridizing to a portion of a tracrRNA or b) being non-covalently bound by an effector protein. In some embodiments, the second sequence is referred to as a repeat sequence. In a dual nucleic acid system, where a crRNA and a tracrRNA forms a complex with an effector protein, a crRNA includes the first sequence that hybridizes to the target sequence of the target nucleic acid and the second sequence hybridizes to a portion of the tracrRNA.

[104] The term, "donor nucleic acid," as used herein, refers to a nucleic acid that is incorporated into a target nucleic acid or target sequence.

[105] The term, "donor nucleotide," as used herein, refers to a single nucleotide that is incorporated into a target nucleic acid. A nucleotide is typically inserted at a site of cleavage by an effector protein.

[106] The term, "dual nucleic acid system," as used herein, refers to a system that uses a transactivated or transactivating tracrRNA-crRNA duplex complexed with one or more polypeptides described herein, wherein the complex is capable of interacting with a target nucleic acid in a sequence selective manner.

[107] The term, "effector protein," as used herein, refers to a protein, polypeptide, or peptide that non-covalently binds to a guide nucleic acid to form a complex that contacts a target nucleic acid, wherein at least a portion of the guide nucleic acid hybridizes to a target sequence of the target nucleic acid. A complex between an effector protein and a guide nucleic acid can include multiple effector proteins or a single effector protein. In some embodiments, the effector protein modifies the target nucleic acid when the complex contacts the target nucleic acid. In some embodiments, the effector protein does not modify the target nucleic acid, but it is fused to a fusion partner protein that modifies the target nucleic acid when the complex contacts the target nucleic acid. A non-limiting example of an effector protein modifying a target nucleic acid is cleaving of a phosphodiester bond of the target nucleic acid. Additional examples of modifications an effector protein can make to target nucleic acids are described herein and throughout.

[108] The term, "functional human protein," as used herein, refers to a protein that retains at least some if not all enzymatic activity relative to the wildtype protein. A functional human protein can also include an human protein having enhanced enzymatic activity relative to the wildtype protein.

[109] The term, "functional fragment," as used herein, refers to a fragment of a protein that retains some function relative to the entire protein. Non-limiting examples of functions are nucleic acid binding, protein binding, nuclease activity, nickase activity, deaminase activity, demethylase activity, or acetylation activity.

[110] The terms, "fusion effector protein," "fusion protein," and "fusion polypeptide," as used herein, refer to a protein comprising at least two heterologous polypeptides. Often a fusion effector protein comprises an effector protein and a fusion partner protein. In general, the fusion partner protein is not an effector protein. Examples of fusion partner proteins are provided herein.

[111] The terms, "fusion partner protein" and "fusion partner," as used herein, refer to a protein, polypeptide or peptide that is fused to an effector protein. The fusion partner generally imparts some function to the fusion protein that is not provided by the effector protein. The fusion partner may modify a target nucleic acid, including changing a nucleobase of the target nucleic acid and making a chemical modification to one or more nucleotides of the target nucleic acid.

[112] The term, "functional domain," as used herein, refers to a region of one or more amino acids in a protein that is required for an activity of the protein, or the full extent of that activity, as measured in an *in vitro* assay. Activities include, but are not limited to nucleic acid binding, nucleic acid modification, nucleic acid cleavage, protein binding. The absence of the functional domain, including mutations of the functional domain, would abolish or reduce activity.

[113] The term, "genetic disease", as used herein, refers to a disease, disorder, condition, or syndrome caused by one or more mutations in the DNA of an organism. Mutations can be due to several different cellular mechanisms, including, but not limited to, an error in DNA replication, recombination, or repair, or due to environmental factors. A genetic disease comprises, in some embodiments, a single gene disorder, a chromosome disorder, or a multifactorial disorder.

[114] The term, "guide nucleic acid," as used herein, refers to a nucleic acid comprising: a first nucleotide sequence that hybridizes to a target nucleic acid; and a second nucleotide sequence that is capable of connecting an effector protein to the nucleic acid by either a) hybridizing to a portion of an additional nucleic acid that is bound by an effector protein (*e.g.*, a tracrRNA) or b) being non-covalently bound by an effector protein. The first sequence may be referred to herein as a spacer sequence. The second sequence may be referred to herein as a repeat sequence or an intermediary sequence. In some embodiments, the first sequence is located 5' of the second nucleotide sequence. In some embodiments, the first sequence is located 3' of the second nucleotide sequence. In some embodiments, the first nucleotide sequence is linked to 5' or 3' end of the second nucleotide sequence. In some embodiments, the first nucleotide sequence is linked to the second nucleotide sequence by a linker nucleic acid. In some embodiments, the linker nucleic acid comprises one, two, three, four or five nucleotide bases. In

some embodiments, the linker nucleic acid comprises a polynucleotide having two, three, four or five nucleotide bases.

[115] The term, "handle sequence," as used herein, in the context of a sgRNA, refers to a portion of the sgRNA that is capable of being non-covalently bound by an effector protein. A handle sequence can also include the portion of the sgRNA that connects the portion of the sgRNA capable of being non-covalently bound by an effector protein to a nucleotide sequence that is hybridizable to a target nucleic acid. A portion of the nucleotide sequence of a handle sequence may contain or be derived from a tracrRNA. For example, in some aspects, a handle sequence can include a portion of a tracrRNA that is capable of being non-covalently bound by an effector protein, but does not include all or a part of the portion of a tracrRNA that hybridizes to a portion of a crRNA as found in a dual nucleic acid system. In some aspects, a handle sequence can include a portion of a tracrRNA as well as a portion of a repeat sequence, which can optionally be connected by a linker. In some aspects, a handle sequence in the context of a sgRNA can also be described as the portion of the sgRNA that does not hybridize to a target sequence in a target nucleic acid (*e.g.*, a spacer sequence).

[116] The term, "heterologous," as used herein, means a nucleotide or polypeptide sequence that is not found in a native nucleic acid or protein, respectively. In some embodiments, fusion proteins comprise an effector protein and a fusion partner protein, wherein the fusion partner protein is heterologous to an effector protein. These fusion proteins may be referred to as a "heterologous protein." A protein that is heterologous to the effector protein is a protein that is not covalently linked via an amide bond to the effector protein in nature. In some embodiments, a heterologous protein is not encoded by a species that encodes the effector protein. In some embodiments, the heterologous protein exhibits an activity (*e.g.*, enzymatic activity) when it is fused to the effector protein. In some embodiments, the heterologous protein exhibits increased or reduced activity (*e.g.*, enzymatic activity) when it is fused to the effector protein, relative to when it is not fused to the effector protein. In some embodiments, the heterologous protein exhibits an activity (*e.g.*, enzymatic activity) that it does not exhibit when it is fused to the effector protein. A guide nucleic acid may comprise a first sequence and a second sequence, wherein the first sequence and the second sequence are not found covalently linked via a phosphodiester bond in nature. Thus, the first sequence is considered to be heterologous with the second sequence, and the guide nucleic acid may be referred to as a heterologous guide nucleic acid.

[117] The terms, "intermediary RNA" and "intermediary sequence," as used herein, in a context of a single nucleic acid system, refers to a nucleotide sequence in a handle sequence, wherein the nucleotide sequence is capable of, at least partially, being non-covalently bound to an effector protein to form a complex (*e.g.,* an RNP complex). An intermediary sequence is not a transactivating nucleic acid in systems, methods, and compositions described herein.

[118] The term, "*in vitro,*" as used herein, is used to describe an event that takes places contained in a container for holding laboratory reagent such that it is separated from the biological source from which the material is obtained. *In vitro* assays can encompass cell-based assays in which living or dead cells

are employed. *In vitro* assays can also encompass a cell-free assay in which no intact cells are employed. The term "*in vivo*" is used to describe an event that takes place in a subject's body. The term "*ex vivo*" is used to describe an event that takes place outside of a subject's body. An *ex vivo* assay is not performed on a subject. Rather, it is performed upon a sample separate from a subject. An example of an *ex vivo* assay performed on a sample is an "*in vitro*" assay.

[119] The term, "linked amino acids" as used herein, refers to at least two amino acids linked by an amide bond.

[120] The term, "linker," as used herein, refers to a bond or molecule that links a first polypeptide to a second polypeptide or a first nucleic acid to a second nucleic acid. A "peptide linker" comprises at least two amino acids linked by an amide bond.

[121] The term, "modified target nucleic acid," as used herein, refers to a target nucleic acid, wherein the target nucleic acid has undergone a modification, for example, after contact with an effector protein. In some cases, the modification is an alteration in the sequence of the target nucleic acid. In some cases, the modified target nucleic acid comprises an insertion, deletion, replacement, or combinations thereof of one or more nucleotides compared to the unmodified target nucleic acid.

[122] The term, "mutation associated with a disease," as used herein, refers to the co-occurrence of a mutation and the phenotype of a disease. The mutation may occur in a gene, wherein transcription or translation products from the gene occur at a significantly abnormal level or in an abnormal form in a cell or subject harboring the mutation as compared to a non-disease control subject not having the mutation.

[123] The terms, "non-naturally occurring" and "engineered," as used herein, are used interchangeably and indicate the involvement of the hand of man. The terms, when referring to a nucleic acid, nucleotide, protein, polypeptide, peptide or amino acid, refer to a nucleic acid, nucleotide, protein, polypeptide, peptide or amino acid that is at least substantially free from at least one other feature with which it is naturally associated in nature and as found in nature, and/or contains a modification (*e.g.*, chemical modification, nucleotide sequence, or amino acid sequence) that is not present in the naturally occurring nucleic acid, nucleotide, protein, polypeptide, peptide, or amino acid. The terms, when referring to a composition or system described herein, refer to a composition or system having at least one component that is not naturally associated with the other components of the composition or system. By way of a non-limiting example, a composition may include an effector protein and a guide nucleic acid that do not naturally occur together. Conversely, and as a non-limiting further clarifying example, an effector protein or guide nucleic acid that is "natural," "naturally-occurring," or "found in nature" includes an effector protein and a guide nucleic acid from a cell or organism that have not been genetically modified by the hand of man.

[124] The term, "nucleic acid expression vector," as used herein, refers to a plasmid that can be used to express a nucleic acid of interest.

[125] The term, "nuclear localization signal," as used herein, refers to an entity (*e.g.*, peptide) that facilitates localization of a nucleic acid, protein, or small molecule to the nucleus, when present in a cell that contains a nuclear compartment.

[126] The term, "nuclease activity," as used herein, refers to the enzymatic activity of an enzyme which allows the enzyme to cleave the phosphodiester bonds between the nucleotide subunits of nucleic acids; the term "endonuclease activity" refers to the enzymatic activity of an enzyme which allows the enzyme to cleave the phosphodiester bond within a polynucleotide chain. An enzyme with nuclease activity may be referred to as a "nuclease."

[127] The term, "pharmaceutically acceptable excipient, carrier or diluent," as used herein, refers to any substance formulated alongside the active ingredient of a pharmaceutical composition that allows the active ingredient to retain biological activity and is non-reactive with the subject's immune system. Such a substance can be included for the purpose of long-term stabilization, bulking up solid formulations that contain potent active ingredients in small amounts, or to confer a therapeutic enhancement on the active ingredient in the final dosage form, such as facilitating absorption, reducing viscosity, or enhancing solubility. The selection of appropriate substance can depend upon the route of administration and the dosage form, as well as the active ingredient and other factors. Compositions having such substances can be formulated by well-known conventional methods (*see, e.g.*, Remington's Pharmaceutical Sciences, 18th edition, A. Gennaro, ed., Mack Publishing Co., Easton, Pa., 1990; and Remington, The Science and Practice of Pharmacy 21st Ed. Mack Publishing, 2005).

[128] The term, "protospacer adjacent motif (PAM)," as used herein, refers to a nucleotide sequence found in a target nucleic acid that directs an effector protein to modify the target nucleic acid at a specific location. A PAM sequence may be required for a complex having an effector protein and a guide nucleic acid to hybridize to and modify the target nucleic acid. However, a given effector protein may not require a PAM sequence being present in a target nucleic acid for the effector protein to modify the target nucleic acid.

[129] The term, "recombinant," as used herein, as applied to proteins, polypeptides, peptides and nucleic acids, refers to proteins, polypeptides, peptides and nucleic acids that are products of various combinations of cloning, restriction, and/or ligation steps resulting in a construct having a structural coding or non-coding sequence distinguishable from endogenous nucleic acids found in natural systems. Generally, DNA sequences encoding the structural coding sequence can be assembled from cDNA fragments and short oligonucleotide linkers, or from a series of synthetic oligonucleotides, to provide a synthetic nucleic acid which is capable of being expressed from a recombinant transcriptional unit contained in a cell or in a cell-free transcription and translation system. Such sequences can be provided in the form of an open reading frame uninterrupted by internal non translated sequences, or introns, which are typically present in eukaryotic genes. Genomic DNA comprising the relevant sequences can also be used in the formation of a recombinant gene or transcriptional unit. Sequences of non-translated DNA may be present 5' or 3' from the open reading frame, where such sequences do not

interfere with manipulation or expression of the coding regions and may act to modulate production of a desired product by various mechanisms. Thus, for example, the term "recombinant polynucleotide" or "recombinant nucleic acid" refers to one which is not naturally occurring, *e.g.*, is made by the artificial combination of two otherwise separated segments of sequence through human intervention. This artificial combination is often accomplished by either chemical synthesis means, or by the artificial manipulation of isolated segments of nucleic acids, *e.g.*, by genetic engineering techniques. Such is usually done to replace a codon with a redundant codon encoding the same or a conservative amino acid, while typically introducing or removing a sequence recognition site. Alternatively, it is performed to join together nucleic acid segments of desired functions to generate a desired combination of functions. Similarly, the term "recombinant polypeptide" or "recombinant protein" refers to one which is not naturally occurring, *e.g.*, is made by the artificial combination of two otherwise separated segments of amino sequences through human intervention. Thus, for example, a polypeptide that includes a heterologous amino acid sequence is a recombinant polypeptide.

[130] In some embodiments, the term, "region" as used herein may be used to describe a portion of or all of a corresponding sequence, for example, a spacer region is understood to comprise a portion of or all of a spacer sequence.

[131] The term, "sample," as used herein, generally refers to something comprising a target nucleic acid. In some embodiments, the sample is a biological sample, such as a biological fluid or tissue sample. In some embodiments, the sample is an environmental sample. The sample may be a biological sample or environmental sample that is modified or manipulated. By way of non-limiting example, samples may be modified or manipulated with purification techniques, heat, nucleic acid amplification, salts and buffers.

[132] The term, "single nucleic acid system," as used herein, refers to a system that uses a guide nucleic acid complexed with one or more polypeptides described herein, wherein the complex is capable of interacting with a target nucleic acid in a sequence specific manner, and wherein the guide nucleic acid is capable of non-covalently interacting with the one or more polypeptides described herein, and wherein the guide nucleic acid is capable of hybridizing with a target sequence of the target nucleic acid. A single nucleic acid system lacks a duplex of a guide nucleic acid as hybridized to a second nucleic acid, wherein in such a duplex the second nucleic acid, and not the guide nucleic acid, is capable of interacting with the effector protein. In a single nucleic system, the guide nucleic acid is not transactivating or transactivated. In a single nucleic acid system, the guide nucleic acid-polypeptide complex (*e.g.*, an RNP complex) is not transactivated or transactivating.

[133] The term, "subject," as used herein, refers to a biological entity containing expressed genetic materials. The biological entity can be a plant, animal, or microorganism, including, for example, bacteria, viruses, fungi, and protozoa. The subject can be tissues, cells and their progeny of a biological entity obtained *in vivo* or cultured *in vitro*. The subject can be a mammal. The mammal can be a human.

The subject may be diagnosed or suspected of being at high risk for a disease. In some embodiments, the subject is not necessarily diagnosed or suspected of being at high risk for the disease.

[134] The term, "syndrome", as used herein, refers to a group of symptoms which, taken together, characterize a condition.

[135] The term, "target nucleic acid," as used herein, refers to a nucleic acid that is selected as the nucleic acid for modification, binding, hybridization or any other activity of or interaction with a nucleic acid, protein, polypeptide, or peptide described herein. A target nucleic acid may comprise RNA, DNA, or a combination thereof. A target nucleic acid may be single-stranded (*e.g.*, single-stranded RNA or single-stranded DNA) or double-stranded (*e.g.*, double-stranded DNA).

[136] The term, "target sequence," as used herein, when used in reference to a target nucleic acid, refers to a sequence of nucleotides found within a target nucleic acid. Such a sequence of nucleotides can, for example, hybridize to an equal length portion of a guide nucleic acid. Hybridization of the guide nucleic acid to the target sequence may bring an effector protein into contact with the target nucleic acid.

[137] The term, "target nucleic acid sequence" in some contexts refers to a "target sequence" and/or a "target nucleic acid."

[138] The terms, "transactivating", "trans-activating", "trans-activated", "transactivated" and grammatical equivalents thereof, as used herein, in the context of a dual nucleic acid system refers to an outcome of the system, wherein a polypeptide is enabled to have a binding and/or nuclease activity on a target nucleic acid, by a tracrRNA or a tracrRNA-crRNA duplex.

[139] The term "*trans* cleavage," is used herein, in reference to cleavage (hydrolysis of a phosphodiester bond) of one or more nucleic acids by an effector protein that is complexed with a guide nucleic acid and a target nucleic acid. The one or more nucleic acids may include the target nucleic acid as well as non-target nucleic acids.

[140] The term "trans-activating RNA (tracrRNA)," as used herein, refers to a nucleic acid that comprises a first sequence that is capable of being non-covalently bound by an effector protein. TracrRNAs may comprise a second sequence that hybridizes to a portion of a crRNA, which may be referred to as a repeat hybridization sequence. In some embodiments, tracrRNAs are covalently linked to a crRNA.

[141] The terms "treatment" or "treating," as used herein, are used in reference to a pharmaceutical or other intervention regimen for obtaining beneficial or desired results in the recipient. Beneficial or desired results include but are not limited to a therapeutic benefit and/or a prophylactic benefit. A therapeutic benefit may refer to eradication or amelioration of symptoms or of an underlying disorder being treated. Also, a therapeutic benefit can be achieved with the eradication or amelioration of one or more of the physiological symptoms associated with the underlying disorder such that an improvement is observed in the subject, notwithstanding that the subject may still be afflicted with the underlying disorder. A prophylactic effect includes delaying, preventing, or eliminating the appearance of a disease

or condition, delaying, or eliminating the onset of symptoms of a disease or condition, slowing, halting, or reversing the progression of a disease or condition, or any combination thereof. For prophylactic benefit, a subject at risk of developing a particular disease, or to a subject reporting one or more of the physiological symptoms of a disease may undergo treatment, even though a diagnosis of this disease may not have been made.

[142] The term "viral vector," as used herein, refers to a nucleic acid to be delivered into a host cell via a recombinantly produced virus or viral particle. The nucleic acid may be single-stranded or double stranded, linear or circular, segmented or non-segmented. The nucleic acid may comprise DNA, RNA, or a combination thereof. Non-limiting examples of viruses or viral particles that can deliver a viral vector include retroviruses (*e.g.*, lentiviruses and γ-retroviruses), adenoviruses, arenaviruses, alphaviruses, adeno-associated viruses (AAVs), baculoviruses, vaccinia viruses, herpes simplex viruses and poxviruses. A viral vector delivered by such viruses or viral particles may be referred to by the type of virus to deliver the viral vector (*e.g.*, an AAV viral vector is a viral vector that is to be delivered by an adeno-associated virus). A viral vector referred to by the type of virus to be delivered by the viral vector can contain viral elements (*e.g.*, nucleotide sequences) necessary for packaging of the viral vector into the virus or viral particle, replicating the virus, or other desired viral activities. A virus containing a viral vector may be replication competent, replication deficient or replication defective.

Introduction

[143] Disclosed herein are compositions, systems, and methods comprising at least one of:

a) a polypeptide or a nucleic acid encoding the polypeptide; and

b) a guide nucleic acid or a nucleic acid encoding the guide nucleic acid.

[144] In some embodiments, compositions, systems, and methods comprise a polypeptide or nucleic acid encoding the polypeptide, wherein the polypeptide is an effector protein, also referred to as a programmable nuclease or programmable nickase. Effector proteins and programmable nucleases/nickases are described herein and throughout. In general, programmable nucleases (*e.g.*, effector protein) are proteins that bind and cleave nucleic acids in a sequence specific manner. A programmable nuclease (*e.g.*, effector protein) may bind a target region of a nucleic acid and cleave the nucleic acid within the target region or a position adjacent to the target region. In some embodiments, a programmable nuclease (*e.g.*, effector protein) is activated when it binds a target region of a nucleic acid to cleave regions of the nucleic acid that are near, but not adjacent to the target region. A programmable nuclease (*e.g.*, effector protein) may be coupled to a guide nucleic acid (e.g., crRNA or sgRNA) that imparts activity or sequence selectivity to the programmable nuclease. In some cases, a composition comprising effector proteins and guide nucleic acids further comprise a trans-activating crRNA (tracrRNA), at least a portion of which interacts with the programmable nuclease (*e.g.*, effector protein). In some cases, a tracrRNA is provided separately from the guide nucleic acid. The tracrRNA may hybridize to a portion of the guide nucleic acid that does not hybridize to the target nucleic acid.

Programmable nucleases (*e.g.,* effector proteins) may cleave nucleic acids, including single stranded RNA (ssRNA), double stranded DNA (dsDNA), and single-stranded DNA (ssDNA). Programmable nucleases (*e.g.,* effector proteins) may provide *cis* cleavage activity, *trans* cleavage activity, nickase activity, or a combination thereof. *Cis* cleavage activity is cleavage of a target nucleic acid that is hybridized to a guide RNA (crRNA or sgRNA), wherein cleavage occurs within or directly adjacent to the region of the target nucleic acid that is hybridized to guide RNA. *Trans* cleavage activity (also referred to as transcollateral cleavage) is cleavage of ssDNA or ssRNA that is near, but not hybridized to the guide RNA. *Trans c*leavage activity is triggered by the hybridization of guide RNA to the target nucleic acid. Nickase activity is the selective cleavage of one strand of a dsDNA molecule.

[145] Programmable CRISPR-associated (Cas) nucleases, through their ability to cleave DNA at a precise target location in the genome of a wide variety of cells and organisms, allow for precise and efficient editing of DNA sequences of interest. SSBs and DSBs are an effective way to disrupt a gene of interest, generate DNA or RNA modifications, and to treat genetic disease through gene correction.

[146] Disclosed herein are non-naturally occurring compositions and systems comprising at least one of an engineered effector protein and an engineered guide nucleic acid, which may simply be referred to herein as an effector protein and a guide nucleic acid, respectively. In some embodiments, compositions, systems and methods comprise an effector protein or a use thereof. In some embodiments, compositions, systems and methods comprise an isolated effector protein or a use thereof. In general, an effector protein and a guide nucleic acid refer to an effector protein and a guide nucleic acid, respectively, that are not found in nature. In some embodiments, systems and compositions herein comprise at least one non-naturally occurring component. For example, compositions and systems may comprise a guide nucleic acid, wherein the sequence of the guide nucleic acid is different or modified from that of a naturally-occurring guide nucleic acid.

[147] In some embodiments, compositions, systems, and methods comprise at least two components that do not naturally occur together. For example, disclosed compositions, systems, and methods comprise a guide nucleic acid comprising a repeat region and a spacer region which do not naturally occur together. Also, by way of example, composition and systems may comprise a guide nucleic acid and an effector protein that do not naturally occur together. Likewise, by way of non-limiting example, disclosed compositions, systems, and methods may comprise a ribonucleotide-protein (RNP) complex comprising an effector protein and a guide nucleic acid that do not occur together in nature. Conversely, and for clarity, an effector protein or guide nucleic acid that is "natural," "naturally-occurring," or "found in nature" includes effector proteins and guide nucleic acids from cells or organisms that have not been genetically modified by a human or machine.

[148] In some embodiments, the guide nucleic acid comprises a non-natural nucleotide sequence. In some embodiments, the non-natural nucleotide sequence is a nucleotide sequence that is not found in nature. The non-natural nucleotide sequence may comprise a portion of a naturally-occurring sequence, wherein the portion of the naturally-occurring sequence is not present in nature absent the remainder of

the naturally-occurring sequence. In some embodiments, the non-naturally occurring sequence is generated by either chemical synthesis means, or by the artificial manipulation of isolated segments of nucleic acids, *e.g.,* by genetic engineering techniques. In some embodiments, the guide nucleic acid comprises two naturally-occurring sequences arranged in an order or proximity that is not observed in nature. In some embodiments, compositions and systems comprise a ribonucleotide complex comprising an effector protein and a guide nucleic acid that do not occur together in nature. In some embodiments, compositions and systems comprise at least two components that do not occur together in nature, wherein the at least two components comprise at least one of an effector protein, a fusion partner and a guide nucleic acid. Guide nucleic acids may comprise a first sequence and a second sequence that do not occur naturally together. For example, a guide nucleic acid may comprise a sequence of a naturally-occurring repeat region and a spacer region that is complementary to a naturally-occurring eukaryotic sequence. The guide nucleic acid may comprise a sequence of a repeat region that occurs naturally in an organism and a spacer region that does not occur naturally in that organism. A guide nucleic acid may comprise a first sequence that occurs in a first organism and a second sequence that occurs in a second organism, wherein the first organism and the second organism are different. The guide nucleic acid may comprise a third sequence disposed at a 3' or 5' end of the guide nucleic acid, or between the first and second sequences of the guide nucleic acid. For example, a guide nucleic acid may comprise a naturally occurring crRNA and tracrRNA sequence coupled by a linker sequence. In some embodiments, the guide nucleic acid comprises two heterologous sequences arranged in an order or proximity that is not observed in nature. Therefore, compositions described herein are not naturally occurring.

[149] In some embodiments, compositions and, systems, and methods described herein comprise an effector protein that is similar to a naturally occurring effector protein. The effector protein may lack a portion of the naturally occurring effector protein. The effector protein may comprise a mutation relative to the naturally-occurring effector protein, wherein the mutation is not found in nature. The effector protein may also comprise at least one additional amino acid relative to the naturally-occurring effector protein . For example, the effector protein may comprise an addition of a nuclear localization signal relative to the natural occurring effector protein. In some embodiments, a nucleotide sequence encoding the effector protein is codon optimized (*e.g.,* for expression, in a eukaryotic cell) relative to the naturally occurring sequence.

I. Polypeptide Systems

[150] Provided herein are compositions, systems and methods comprising a polypeptide or polypeptide system, wherein the polypeptide or polypeptide system described herein comprises one or more effector proteins or variants thereof, one or more effector partners or variants thereof, one or more linkers for peptides, or combinations thereof. In some embodiments, a variant is a form or version of a protein that differs from a naturally occurring protein or wildtype protein. For example, a variant may have one or more amino acid substitutions, insertions, or deletions relative to a wildtype protein. In

some embodiments, a variant is at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, or at least 99% identical to a respective wildtype protein. A variant may have a different function or activity relative to the naturally occurring or wildtype protein.

[151] A polypeptide may include coded and non-coded amino acids, chemically or biochemically modified or derivatized amino acids, and polypeptides having modified peptide backbones. Accordingly, polypeptides as described herein may comprise one or more mutations, one or more engineered modifications, or both, relative to a naturally occurring or wildtype protein. It is understood that when describing coding sequences of polypeptides described herein, said coding sequences do not necessarily require a codon encoding an N-terminal Methionine (M) or a Valine (V) as described for the effector proteins described herein. One skilled in the art would understand that a start codon could be replaced or substituted with a start codon that encodes for an amino acid residue sufficient for initiating translation in a host cell. In some embodiments, when a heterologous peptide, such as a fusion partner protein, protein tag or nuclear localization signal (NLS), is located at the N terminus of the effector protein, a start codon for the heterologous peptide serves as a start codon for the effector protein as well. Thus, the natural start codon encoding an amino acid residue sufficient for initiating translation (*e.g.,* Methionine (M) or a Valine (V)) of the effector protein may be removed or absent.

Effector Proteins

[152] Provided herein are compositions, systems, and methods comprising a polypeptide (*e.g.,* an effector protein, an effector partner, a fusion protein, or a combination thereof). In some embodiments, compositions, systems, and methods described herein comprise an effector protein or a use thereof. In some embodiments, the effector protein is a protein, polypeptide, or peptide that non-covalently binds to a guide nucleic acid to form a complex that interacts with a target nucleic acid.

[153] In general, effector proteins provided herein interact with a guide nucleic acid to form a complex. In some embodiments, the complex interacts with a target nucleic acid. In some embodiments, an interaction between the complex and a target nucleic acid comprises one or more of: recognition of a protospacer adjacent motif (PAM) sequence within the target nucleic acid by the effector protein, hybridization of the guide nucleic acid to the target nucleic acid, modification of the target nucleic acid by the effector protein, or combinations thereof. In some embodiments, recognition of a PAM sequence within a target nucleic acid may direct the modification activity of an effector protein. In some embodiments, recognition of a PAM sequence adjacent to a target sequence of a target nucleic acid may direct the modification activity of an effector protein. Provided herein, in certain embodiments, are compositions that comprise one or more effector proteins.

[154] Modification activity of an effector protein or an engineered protein described herein may be cleavage activity, binding activity, insertion activity, substitution activity, and the like. Modification activity of an effector protein may result in: cleavage of at least one strand of a target nucleic acid, deletion of one or more nucleotides of a target nucleic acid, insertion of one or more nucleotides into

proximity of a target nucleic acid, substitution of one or more nucleotides of a target nucleic acid with an alternative nucleotide, more than one of the foregoing, or any combination thereof. In some embodiments, modification of a target sequence in the target nucleic acid comprises introducing or removing epigenetic modification(s). In some embodiments, an ability of an effector protein to edit a target nucleic acid may depend upon the effector protein being complexed with a guide nucleic acid, the guide nucleic acid being hybridized to a target sequence of the target nucleic acid, the distance between the target sequence and a PAM sequence, or combinations thereof. A target nucleic acid comprises a target strand and a non-target strand. Accordingly, in some embodiments, the effector protein may edit a target strand and/or a non-target strand of a target nucleic acid.

[155] The modification of the target nucleic acid generated by an effector protein may, as a non-limiting example, result in modulation of the expression of the target nucleic acid (*e.g.*, increasing or decreasing expression of the nucleic acid) or modulation of the activity of a translation product of the target nucleic acid (*e.g.*, inactivation of a protein binding to an RNA molecule or hybridization). Accordingly, in some embodiments, provided herein are methods of editing a target nucleic acid using an effector protein of the present disclosure, or compositions or systems thereof. Also provided herein are methods of modulating expression of a target nucleic acid using an effector protein of the present disclosure, or compositions or systems thereof. Further provided herein are methods of modulating the activity of a translation product of a target nucleic acid using an effector protein of the present disclosure, or compositions or systems thereof.

[156] In some embodiments, effector proteins disclosed herein may provide nucleic acid cleavage activity, such as *cis* cleavage activity, *trans* cleavage activity, nickase activity, nuclease activity, or a combination thereof. Effector proteins disclosed herein may cleave nucleic acids, including single stranded RNA (ssRNA), double stranded DNA (dsDNA), and single-stranded DNA (ssDNA).

[157] In some embodiments, effector proteins disclosed herein may provide catalytic activity (*e.g.*, cleavage activity, nickase activity, nuclease activity, other activity, or combinations thereof) similar to that of a naturally-occurring effector protein, such as, for example, a naturally-occurring effector protein with reduced cleavage activity (*e.g.*, Cas 14, CasPhi.12, CasM.265466) including *cis* cleavage activity, *trans* cleavage activity, or combinations thereof. In some embodiments, effector proteins disclosed herein may be fused to effector partners or fusion proteins wherein the effector partners or fusion proteins are capable of some function or activity not provided by an effector protein. An effector protein may be brought into proximity of a target nucleic acid in the presence of a guide nucleic acid when the guide nucleic acid includes a nucleotide sequence that is complementary with a target sequence in the target nucleic acid. The ability of an effector protein to modify a target nucleic acid may be dependent upon the effector protein being bound to a guide nucleic acid and the guide nucleic acid being hybridized to a target nucleic acid. An effector protein may also recognize (*e.g.*, non-covalently interact) a protospacer adjacent motif (PAM) sequence present in the target nucleic acid, which may direct the modification activity of the effector protein. An effector protein may modify a nucleic acid by cis

cleavage or trans cleavage. The modification of the target nucleic acid generated by an effector protein may, as a non-limiting example, result in expression of a protein that is encoded by a donor nucleic acid.

[158] An effector protein may be a CRISPR-associated ("Cas") protein. An effector protein may function as a single protein, including a single protein that is capable of binding to a guide nucleic acid and modifying a target nucleic acid. Alternatively, an effector protein may function as part of a multiprotein complex, including, for example, a complex having two or more effector proteins, including two or more of the same effector proteins (*e.g.*, dimer or multimer). An effector protein, when functioning in a multiprotein complex, may have only one functional activity (*e.g.*, binding to a guide nucleic acid), while other effector proteins present in the multiprotein complex are capable of the other functional activity (*e.g.*, modifying a target nucleic acid). In some embodiments, an effector protein, when functioning in a multiprotein complex, may have differing and/or complementary functional activity to other effector proteins in the multiprotein complex. In such embodiments, the functional activity includes but not limited to substrate selectivity, specificity, and/or affinity. An effector protein may be a modified effector protein having reduced modification activity (*e.g.*, a catalytically defective effector protein) or no modification activity (*e.g.*, a catalytically inactive effector protein). Accordingly, an effector protein as used herein encompasses a modified or programmable nuclease (*e.g.*, effector protein) that does not have nuclease activity. In some embodiments, the complementary functional activity of effector proteins comprising modified or artificial base pairs can be based on other types of hydrogen bonding and/or hydrophobicity of bases and/or shape complementarity between bases. Multimeric complexes, and functions thereof, are described in further detail below. An effector protein may be a modified effector protein having increased modification activity and/or increased substrate binding activity (*e.g.*, substrate selectivity, specificity, and/or affinity). Alternatively, or in addition, an effector protein may be a catalytically inactive effector protein having reduced modification activity or no modification activity.

[159] In some embodiments, effector proteins described herein comprise one or more functional domains. Effector protein functional domains can include a protospacer adjacent motif (PAM)-interacting domain, an oligonucleotide-interacting domain, one or more recognition domains, a non-target strand interacting domain, and a RuvC domain. A PAM interacting domain can be a target strand PAM interacting domain (TPID) or a non-target strand PAM interacting domain (NTPID). In some embodiments, a PAM interacting domain, such as a TPID or a NTPID, on an effector protein describes a region of an effector protein that interacts with target nucleic acid. In some embodiments, the effector proteins comprise a RuvC domain. In some embodiments, the RuvC domain may be defined by a single, contiguous sequence, or a set of RuvC subdomains that are not contiguous with respect to the primary amino acid sequence of the protein. An effector protein of the present disclosure may include multiple RuvC subdomains, which may combine to generate a RuvC domain with substrate binding or catalytic activity. For example, an effector protein may include three RuvC subdomains (RuvC-I, RuvC-II, and

RuvC-III) that are not contiguous with respect to the primary amino acid sequence of the effector protein but form a RuvC domain once the protein is produced and folds. In many embodiments, effector proteins comprise a recognition domain with a binding affinity for a guide nucleic acid or for a guide nucleic acid-target nucleic acid heteroduplex. In some embodiments, the REC domain is an α-helical recognition region or lobe. A CRISPR/Cas protein can contain 2 REC domains (REC1 and REC2) which generally helps to accommodate and stabilize the guide nucleic acid and target nucleic acid hybrid. An effector protein may comprise a zinc finger domain. In some embodiments, the effector protein does not comprise an HNH domain. In some embodiments, a RuvC domain, as used herein, refers to a region of an effector protein that is capable of cleaving a target nucleic acid, and in some embodiments, of processing a pre-crRNA. In some embodiments, the RuvC domain is located near the C-terminus of the effector protein. A single RuvC domain may comprise RuvC subdomains, for example a RuvCI subdomain, a RuvCII subdomain and a RuvCIII subdomain. The term "RuvC" domain can also refer to a "RuvC-like" domain. Various RuvC-like domains are known in the art and are easily identified using online tools such as InterPro (https://www.ebi.ac.uk/interpro/). For example, a RuvC-like domain may be a domain which shares homology with a region of TnpB proteins of the IS605 and other related families of transposons.

[160] **TABLE 1A** and **TABLE 1B** provides an illustrative amino acid sequence of an effector protein. In some embodiments, an effector protein is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to the amino acid sequence as set forth in **TABLE 1A** and **TABLE 1B**.

[161] In some embodiments, compositions, systems and methods described herein comprise an effector protein, or a nucleic acid encoding the effector protein, wherein the effector protein comprises an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identical to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the amino acid sequence of the effector protein is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, an effector protein provided herein comprises an amino acid sequence that is at least 65% identical to the amino acid sequence recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, an effector protein provided herein comprises an amino acid sequence that is at least 70% identical to the sequence recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, an effector protein provided herein comprises an amino acid sequence that is at least 75% identical to the amino acid sequence recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, an effector protein provided herein comprises an amino acid sequence that is at least 80% identical to the amino acid sequence recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, an effector protein provided herein comprises an amino acid sequence that is at least 85% identical to the amino acid sequence recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, an

effector protein provided herein comprises an amino acid sequence that is at least 90% identical to the amino acid sequence recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, an effector protein provided herein comprises an amino acid sequence that is at least 95% identical to the amino acid sequence recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, an effector protein provided herein comprises an amino acid sequence that is at least 98% identical to the amino acid sequence recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, an effector protein provided herein comprises an amino acid sequence that is at least 99% identical to the amino acid sequence recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, an effector protein provided herein comprises an amino acid sequence that is identical to the amino acid sequence recited in **TABLE 1A** and **TABLE 1B**.

[162] In some embodiments, compositions, systems and methods described herein comprise an effector protein, or a nucleic acid encoding the effector protein, wherein the effector protein comprises an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identical to any one of **SEQ ID NO: 1-231**. In some embodiments, the amino acid sequence of the effector protein is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of **SEQ ID NO: 1- 231**.

[163] In some embodiments, compositions, systems and methods described herein comprise an effector protein capable of interacting with a nucleic acid, such as a guide nucleic acid, to form a complex (*e.g.,* a ribonucleotide protein complex (RNP) complex), wherein the complex interacts with a target nucleic acid. In some embodiments, of the complex comprises one or more nucleic acids and one or more polypeptides. The one or more nucleic acids may comprise deoxyribonucleotides (DNA), ribonucleotides (RNA), biochemically or chemically modified nucleobases (*e.g.,* one or more engineered modifications described herein), or combinations thereof.

[164] In some embodiments, compositions, systems and methods described herein comprise an effector protein comprising an amino acid sequence for which the % identity is calculated by dividing the total number of the aligned residues by the number of the residues that are identical between the respective positions of the at least two amino acid sequences and multiplying by 100. In some embodiments, the percent of residues that are identical between respective positions of two amino acid sequences when two amino acid sequences are aligned for maximum sequence identity is referred to as percent identity (*e.g.,* % identity). In some embodiments, the non-covalent interaction between macromolecules (*e.g.,* between two polypeptides, between a polypeptide and a nucleic acid; between a polypeptide/guide nucleic acid complex and a target nucleic acid) is referred to as binding or associating. Non-limiting examples of non-covalent interactions are ionic bonds, hydrogen bonds, van der Waals and hydrophobic interactions. Not all components of a binding interaction need be sequence-specific (*e.g.,* contacts with phosphate residues in a DNA backbone), but some portions of a binding interaction may be sequence-specific. In some embodiments, a nucleotide sequence that is able to

noncovalently interact or anneal to another nucleotide sequence in a sequence-specific, antiparallel, manner (*i.e.,* a nucleotide sequence specifically interacts to a complementary nucleotide sequence) is considered to be hybridizable or capable of hybridization. In some embodiments, the nucleotide sequence is able to noncovalently interact or anneal to another nucleotide sequence, as detailed herein, under the appropriate *in vitro* and/or *in vivo* conditions of temperature and solution ionic strength. In addition, for hybridization between two RNA molecules (*e.g.,* dsRNA), and for hybridization of a DNA molecule with an RNA molecule (*e.g.,* when a DNA target nucleic acid base pairs with a guide RNA, *etc.*): guanine (G) can also base pair with uracil (U). For example, G/U base-pairing is at least partially responsible for the degeneracy (*i.e.,* redundancy) of the genetic code in the context of tRNA anti-codon base-pairing with codons in mRNA. Thus, a guanine (G) can be considered complementary to both an uracil (U) and to an adenine (A). Accordingly, when a G/U base-pair can be made at a given nucleotide position, the position is not considered to be non-complementary, but is instead considered to be complementary. While hybridization typically occurs between two nucleotide sequences that are complementary, mismatches between bases are possible. It is understood that two nucleotide sequences need not be 100% complementary to be specifically hybridizable, hybridizable, partially hybridizable, or for hybridization to occur. Moreover, a nucleotide sequence may hybridize over one or more segments such that intervening or adjacent segments are not involved in the hybridization event (*e.g.,* a bulge, a loop structure or hairpin structure, *etc.*). The conditions appropriate for hybridization between two nucleotide sequences depend on the length of the sequence and the degree of complementarity, variables which are well known in the art. For hybridizations between nucleic acids with short stretches of complementarity (*e.g.,* complementarity over 35 or less, 30 or less, 25 or less, 22 or less, 20 or less, or 18 or less nucleotides) the position of mismatches may become important (see Sambrook *et al.,* supra, 11.7-11.8). Typically, the length for a hybridizable nucleic acid is 8 nucleotides or more (*e.g.,* 10 nucleotides or more, 12 nucleotides or more, 15 nucleotides or more, 20 nucleotides or more, 22 nucleotides or more, 25 nucleotides or more, or 30 nucleotides or more). Any suitable *in vitro* assay may be utilized to assess whether two sequences hybridize. One such assay is a melting point analysis where the greater the degree of complementarity between two nucleotide sequences, the greater the value of the melting temperature (Tm) for hybrids of nucleic acids having those sequences. The conditions of temperature and ionic strength determine the stringency of the hybridization. Temperature, wash solution salt concentration, and other conditions may be adjusted as necessary according to factors such as length of the region of complementation and the degree of complementation. Hybridization and washing conditions are well known and exemplified in Sambrook, J. and Russell, W., Molecular Cloning: A Laboratory Manual, Third Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (2001); and in Green, M. and Sambrook, J., Molecular Cloning: A Laboratory Manual, Fourth Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (2012).

[165] In some embodiments, compositions, systems and methods described herein comprise a nucleic acid encoding the effector protein, wherein the nucleic acid encoding the effector protein is a messenger RNA.

[166] An effector protein may be small, which may be beneficial for nucleic acid detection or editing (for example, the effector protein may be less likely to adsorb to a surface or another biological species due to its small size). The smaller nature of these effector proteins may allow for them to be more easily packaged and delivered with higher efficiency in the context of genome editing and more readily incorporated as a reagent in an assay. In some embodiments, the length of the effector protein is at least 400 linked amino acid residues. In some embodiments, the length of the effector protein is less than 500 linked amino acid residues. In some embodiments, the length of the effector protein is about 400 to about 500 linked amino acid residues. In some embodiments, the length of the effector protein is about 450 to about 550, about 400 to about 420, about 420 to about 440, about 440 to about 460, about 460 to about 480, about 480 to about 500, about 500 to about 520, about 520 to about 540, about 540 to about 560, about 560 to about 580, about 580 to about 600, about 600 to about 620, about 620 to about 640, about 640 to about 660, about 660 to about 680, about 680 to about 700 linked amino acids.

[167] In some embodiments, the amino acid sequence of an effector protein provided herein comprises at least about 200, at least about 220, at least about 240, at least about 260, at least about 280, at least about 300, at least about 320, at least about 340, at least about 360, at least about 380, at least about 400 contiguous amino acids, at least about 420 contiguous amino acids, at least about 440 contiguous amino acids, at least about 460 contiguous amino acids, at least about 480 contiguous amino acids, at least about 500 contiguous amino acids, at least about 520 contiguous amino acids, at least about 540 contiguous amino acids, at least about 560 contiguous amino acids, at least about 580 contiguous amino acids, at least about 600 contiguous amino acids, at least about 620 contiguous amino acids, at least about 640 contiguous amino acids, at least about 660 contiguous amino acids, at least about 680 contiguous amino acids, at least about 700 contiguous amino acids, or more of the amino acid sequence recited in **TABLE 1A** and **TABLE 1B**. In certain embodiments, compositions comprise an effector protein, or a nucleic acid encoding the effector protein, wherein the amino acid sequence of the effector protein comprises at least about 200, at least about 220, at least about 240, at least about 260, at least about 280, at least about 300, at least about 320, at least about 340, at least about 360, at least about 380, or at least about 400 contiguous amino acids or more of any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In certain embodiments, compositions comprise an effector protein, or a nucleic acid encoding the effector protein, wherein the amino acid sequence of the effector protein comprises at least about 200 contiguous amino acids or more of any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In certain embodiments, compositions comprise an effector protein, or a nucleic acid encoding the effector protein, wherein the amino acid sequence of the effector protein comprises at least about 300 contiguous amino acids or more of any one of the amino acid sequences recited **TABLE 1A** and **TABLE 1B**. In certain embodiments, compositions

comprise an effector protein, or a nucleic acid encoding the effector protein, wherein the amino acid sequence of the effector protein comprises at least about 400 contiguous amino acids or more of any one of the amino acid sequences recited **TABLE 1A** and **TABLE 1B**.

[168] In some embodiments, an effector protein may be able to recognize a variety of PAMs as described herein. In some embodiments, effector proteins described herein may provide blunt or short stagger ends. Blunt cutting may be advantageous over the staggered cutting that is provided by other effector proteins, as there is a less likely chance of spontaneous (also referred to as perfect) repair which may decrease the chances of successful target nucleic acid editing and/or donor nucleic acid insertion.

[169] In some embodiments, compositions, systems, and methods described herein comprise an effector protein or a nucleic acid encoding the effector protein, wherein the effector protein comprises one or more amino acid alterations relative to the amino acid sequence recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the effector protein comprising one or more amino acid alterations is a variant of an effector protein described herein. It is understood that any reference to an effector protein herein also refers to an effector protein variant as described herein. In some embodiments, an amino acid alteration comprises a deletion of an amino acid. In some embodiments, an amino acid alteration comprises an insertion of an amino acid. In some embodiments, an amino acid alteration comprises a conservative amino acid substitution. In some embodiments, an amino acid alteration comprises a non-conservative amino acid substitution. In some embodiments, one or more amino acid alterations comprises a combination of one or more conservative amino acid substitutions and one or more non-conservative amino acid substitutions. When describing a conservative alteration (*e.g.*, conservative substitution) herein, reference is made to the replacement of one amino acid for another such that the replacement takes place within a family of amino acids that are related in their side chains. Conversely, when describing a non-conservative alteration (*e.g.*, non-conservative substitution), reference is made to the replacement of one amino acid residue for another that does not have a related side chain. It is understood that genetically encoded amino acids can be divided into four families having related side chains: (1) acidic (negatively charged): Asp (D), Glu (E); (2) basic (positively charged): Lys (K), Arg (R), His (H); (3) non-polar (hydrophobic): Cys (C), Ala (A), Val (V), Leu (L), Ile (I), Pro (P), Phe (F), Met (M), Trp (W), Gly (G), Tyr (Y), with non-polar also being subdivided into: (i) strongly hydrophobic: Ala (A), Val (V), Leu (L), Ile (I), Met (M), Phe (F); and (ii) moderately hydrophobic: Gly (G), Pro (P), Cys (C), Tyr (Y), Trp (W); and (4) uncharged polar: Asn (N), Gln (Q), Ser (S), Thr (T). Amino acids may be related by aliphatic side chains: Gly (G), Ala (A), Val (V), Leu (L), Ile (I), Ser (S), Thr (T), with Ser (S) and Thr (T) optionally being grouped separately as aliphatic-hydroxyl. Amino acids may be related by aromatic side chains: Phe (F), Tyr (Y), Trp (W). Amino acids may be related by amide side chains: Asn (N), Gln (Q). Amino acids may be related by sulfur-containing side chains: Cys (C) and Met (M). In some embodiments, an effector protein or a nucleic acid encoding the effector protein comprises one or more alterations or more relative to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B.** In some embodiments, an effector protein or a nucleic acid

encoding the effector protein comprises 1 amino acid alteration, 2 amino acid alterations, 3 amino acid alterations, 4 amino acid alterations, 5 amino acid alterations, 6 amino acid alterations, 7 amino acid alterations, 8 amino acid alterations, 9 amino acid alterations, 10 amino acid alterations or more relative to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B.** In some embodiments, the one or more alterations comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, at least twelve, at least sixteen, at least twenty, or more amino acid alterations relative to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the one or more alterations comprises one to twenty, one to sixteen, one to twelve, one to eight, one to four, four to twenty, four to sixteen, four to twelve, four to eight, eight to twenty, eight to sixteen, eight to twelve, twelve to twenty, twelve to sixteen, or sixteen to twenty amino acid alterations relative to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the one or more alterations comprises one, two, three, four, five, six, seven, eight, nine, ten, or more amino acid alterations relative to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**.

[170] In some embodiments, compositions, systems, and methods described herein comprise an effector protein, or a nucleic acid encoding the effector protein, wherein the effector protein comprises one or more substitutions relative to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the one or more substitutions comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, at least twelve, at least sixteen, at least twenty, or more substitutions relative to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the one or more substitutions comprises one to twenty, one to sixteen, one to twelve, one to eight, one to four, four to twenty, four to sixteen, four to twelve, four to eight, eight to twenty, eight to sixteen, eight to twelve, twelve to twenty, twelve to sixteen, or sixteen to twenty substitutions relative to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the one or more substitutions comprise one, two, three, four, five, six, seven, eight, nine, ten or more substitutions relative to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the one or more substitutions comprise one or more conservative substitutions, one or more non-conservative substitutions, or combinations thereof.

[171] In some embodiments, compositions, systems, and methods described herein comprise an effector protein, or a nucleic acid encoding the effector protein, wherein the effector protein comprises one or more conservative substitutions relative to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the one or more conservative substitutions comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, at least twelve, at least sixteen, at least twenty, or more conservative substitutions relative to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the one or more conservative substitutions comprises one to twenty,

one to sixteen, one to twelve, one to eight, one to four, four to twenty, four to sixteen, four to twelve, four to eight, eight to twenty, eight to sixteen, eight to twelve, twelve to twenty, twelve to sixteen, or sixteen to twenty conservative substitutions relative to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the one or more conservative substitutions comprise one, two, three, four, five, six, seven, eight, nine, ten or more conservative substitutions relative to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**.

[172] In some embodiments, compositions, systems, and methods described herein comprise an effector protein, or a nucleic acid encoding the effector protein, wherein the effector protein comprises one or more non-conservative substitutions relative to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the one or more non-conservative substitutions comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, at least twelve, at least sixteen, at least twenty, or more non-conservative substitutions relative to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the one or more non-conservative substitutions comprises one to twenty, one to sixteen, one to twelve, one to eight, one to four, four to twenty, four to sixteen, four to twelve, four to eight, eight to twenty, eight to sixteen, eight to twelve, twelve to twenty, twelve to sixteen, or sixteen to twenty non-conservative substitutions relative to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the one or more non-conservative substitutions comprise one, two, three, four, five, six, seven, eight, nine, ten or more non-conservative substitutions relative to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**.

[173] In some embodiments, the one or more amino acid alterations may result in a change in activity of the effector protein relative to a naturally-occurring counterpart. For example, and as described in further detail below, the one or more amino acid alteration increases or decreases catalytic activity of the effector protein relative to a naturally-occurring counterpart. In another example, the one or more amino acid alteration increases or decreases binding activity of the effector protein relative to a naturally-occurring counterpart. In some embodiments, the one or more amino acid alterations results in a catalytically inactive effector protein variant.

[174] In some embodiments, the one or more amino acid alterations may result in a change in activity of the effector protein relative to a naturally-occurring counterpart. For example, and as described in further detail below, the one or more amino acid alteration increases or decreases catalytic activity of the effector protein relative to a naturally-occurring counterpart. In some embodiments, the one or more amino acid alterations results in a catalytically inactive effector protein variant.

[175] In some embodiments, a variant effector protein as described herein can carry out a similar enzymatic reaction as corresponding reference effector protein, wherein the reference effector protein comprises any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. It is further understood that the variants of the reference effector protein can include alterations that provide a

beneficial characteristic to effector proteins described herein, including but not limited to, increased activity (*e.g.*, indel activity, catalytic activity, specificity or selectivity and/or affinity for a substrate, such as a target nucleic acid and/or a guide nucleic acid). In some embodiments, the variant can exhibit an activity that is at least the same or higher than the corresponding reference effector protein, that is, it has one or more activities that are the same or higher than the corresponding reference effector protein . For example, variants can have one or more activity that is at least 10%, 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 110%, at least 120%, at least 130%, at least 140%, at least 150%, at least 160%, at least 170%, at least 180%, at least 190%, or at least 200% higher over the corresponding reference effector protein. In some embodiments, activity of variants can be measured relative to the corresponding reference effector protein in a cleavage assay.

[176] In some embodiments, effector proteins provided herein are a variant of any one of the effector proteins of **TABLE 1A** and **TABLE 1B**, wherein the variant comprises one or more amino acid alterations in one or more regions that interact with a substrate such as a target nucleic acid, an engineered guide nucleic acid or a guide nucleic acid-target nucleic acid heteroduplex. In some embodiments, the variant comprises one or more amino acid alterations in a region of the effector protein that affects a substrate binding activity, a catalytic activity, and/or a binding affinity for a substrate such as a target nucleic acid, an engineered guide nucleic acid or a guide nucleic acid-target nucleic acid heteroduplex. In some embodiments, the variant comprises one or more amino acid alterations in a RuvC domain, a REC domain, TPID, NTPID, or combinations thereof. In some embodiments, the one or more alterations are one or more substitutions with Lys (K), Arg (R), or His (H). In some embodiments, variants comprise at least one alteration relative to any one of the amino acid sequences of **TABLE 1A** and **TABLE 1B**, wherein the at least one alteration is a substitution with Lys (K), Arg (R), or His (H).

[177] In some embodiments, compositions, systems, and methods described herein comprise an effector protein or a nucleic acid encoding the effector protein, wherein the amino acid sequence of the effector protein is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identical to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, compositions, systems, and methods described herein comprise an effector protein or a nucleic acid encoding the effector protein, wherein the amino acid sequence of the effector protein is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% similar to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the effector protein comprises one or more amino acid substitutions independently selected from K58W, I80K, N193K, S209F, A218K, E225K, N286K, M295W, M298L, A306K, Y315M or a combination thereof, wherein the amino acid sequence of the effector protein is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO:**

231. In some embodiments, the effector protein comprises one or more amino acid substitutions independently selected from K58W, I80K, N193K, S209F, A218K, E225K, N286K, M295W, M298L, A306K, Y315M or a combination thereof, wherein the amino acid sequence of the effector protein is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% similar to **SEQ ID NO: 231**.

[178] In certain embodiments, compositions, systems, and methods described herein comprise an effector protein, or a nucleic acid encoding the effector protein, wherein the effector protein comprises an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% similar to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. An amino acid sequence of the effector protein is similar to the reference amino acid sequence, when a value that is calculated by dividing a similarity score by the length of the alignment. The similarity of two amino acid sequences can be calculated by using a BLOSUM62 similarity matrix (Henikoff and Henikoff, *Proc. Natl. Acad. Sci. USA.*, 89:10915–10919 (1992)) that is transformed so that any value ≥ 1 is replaced with +1 and any value ≤ 0 is replaced with 0. For example, an Ile (I) to Leu (L) substitution is scored at +2.0 by the BLOSUM62 similarity matrix, which in the transformed matrix is scored at +1. This transformation allows the calculation of percent similarity, rather than a similarity score. Alternately, when comparing two full protein sequences, the proteins can be aligned using pairwise MUSCLE alignment. Then, the % similarity can be scored at each residue and divided by the length of the alignment. For determining % similarity over a protein domain or motif, a multilevel consensus sequence (or PROSITE motif sequence) can be used to identify how strongly each domain or motif is conserved. In calculating the similarity of a domain or motif, the second and third levels of the multilevel sequence are treated as equivalent to the top level. Additionally, if a substitution could be treated as conservative with any of the amino acids in that position of the multilevel consensus sequence, +1 point is assigned. For example, given the multilevel consensus sequence: RLG and YCK, the test sequence QIQ would receive three points. This is because in the transformed BLOSUM62 matrix, each combination is scored as: Q-R: +1; Q-Y: +0; I-L: +1; I-C: +0; Q-G: +0; Q-K: +1 For each position, the highest score is used when calculating similarity. The % similarity can also be calculated using commercially available programs, such as the Geneious Prime software given the parameters matrix = BLOSUM62 and threshold ≥ 1. Accordingly, % similarity represents a value that is calculated by dividing a similarity score by the length of the alignment.

[179] In some embodiments, effector proteins provided herein comprise one or more amino acid substitutions relative to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**, wherein the amino acid sequence of the effector protein is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identical to the corresponding reference amino acid sequence of **TABLE 1A** and **TABLE 1B**. In some embodiments, effector proteins provided herein comprise one or more amino acid substitutions relative to any one of

the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**, wherein the amino acid sequence of the effector protein is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% similar to the corresponding reference amino acid sequence of **TABLE 1A** and **TABLE 1B**. In some embodiments, the one or more amino acid substitutions independently comprise one or more conservative substitutions, one or more non-conservative substitutions, or combinations thereof. In some embodiments, the one or more amino acid substitutions comprise substitution with one or more positively charged amino acid residues. In some embodiments, the positively charged amino acid residue is independently selected from Lys (K), Arg (R), and His (H). Accordingly, in some embodiments, the one or more amino acid substitutions comprise substitutions with Lys (K), Arg (R), His (H), or combinations thereof.

[180] In some embodiments, the effector proteins described herein comprises a substitution of at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, or at least ten amino acids with positively charged amino acids. In some embodiments, the effector proteins described herein comprises a substitution of one, two, three, four, five, six, seven, eight, nine, or ten amino acids with positively charged amino acids. In some embodiments, the effector protein comprises one or more amino acid substitutions independently selected from E109R, H208R, K184R, K38R, L182R, Q183R, S108R, S198R, T114R or a combination thereof, wherein the amino acid sequence of the effector protein is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identical to any one of **SEQ ID NO: 6,** and **228-230**. In some embodiments, the effector protein comprises one or more amino acid substitutions independently selected from E109R, H208R, K184R, K38R, L182R, Q183R, S108R, S198R, T114R or a combination thereof, wherein the amino acid sequence of the effector protein is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% similar to any one of **SEQ ID NO: 6,** and **228-230**. In some embodiments, the effector protein comprises one or more amino acid substitutions independently selected from I80R, T84R, K105R, G210R, C202R, A218R, D220R, E225R, C246R, Q360R or a combination thereof, wherein the amino acid sequence of the effector protein is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 231**. In some embodiments, the effector protein comprises one or more amino acid substitutions independently selected from I80R, T84R, K105R, G210R, C202R, A218R, D220R, E225R, C246R, Q360R or a combination thereof, wherein the amino acid sequence of the effector protein is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% similar to **SEQ ID NO: 231**. In some embodiments, the effector protein comprises one or more amino acid substitutions, wherein the amino acid sequence of the effector protein is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 47**. In some embodiments, the effector protein comprises one or more amino acid substitutions, wherein the amino acid sequence of the effector protein is at least

65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% similar to **SEQ ID NO: 47**.

[181] In some embodiments, the effector protein comprises an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 228**, wherein the effector protein has Arg (R) amino acid at position 26. In some embodiments, the effector protein comprises an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% similar to **SEQ ID NO: 228**, wherein the effector protein has Arg (R) amino acid at position 26.

[182] In some embodiments, the effector protein comprises an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 229**, wherein the effector protein has His (H) amino acid at position 26. In some embodiments, the effector protein comprises an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% similar to **SEQ ID NO: 229**, wherein has His (H) amino acid at position 26.

[183] In some embodiments, the effector protein comprises an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 230**, wherein the effector protein has Lys (K) amino acid at position 26. In some embodiments, the effector protein comprises an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% similar to **SEQ ID NO: 230**, wherein the effector protein has Lys (K) amino acid at position 26.

[184] In some embodiments, the effector proteins described herein comprising one or more amino acid substitutions comprise substitution with one or more positively charged amino acid residues, wherein, the positively charged amino acid residue is independently selected from Lys (K), Arg (R), and His (H), and wherein the amino acid sequence of the effector protein is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identical to the corresponding reference amino acid sequence of **TABLE 1A** and **TABLE 1B**. In some embodiments, the effector protein comprises one or more amino acid substitutions independently selected from E109R, H208R, K184R, K38R, L182R, Q183R, S108R, S198R, T114R or a combination thereof, wherein the amino acid sequence of the effector protein is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identical to any one of **SEQ ID NO: 6,** and **228-230**. In some embodiments, the effector protein comprises one or more amino acid substitutions independently selected from I80R, T84R, K105R, G210R, C202R, A218R, D220R, E225R, C246R, Q360R or a combination thereof, wherein the amino acid sequence of

the effector protein is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 231**.

[185] In some embodiments, the one or more amino acid alterations may result in a change in activity of the effector protein relative to a WT counterpart. For example, and as described in further detail below, the one or more amino acid alteration increases or decreases catalytic activity of the effector protein relative to a WT counterpart. In some embodiments, the one or more amino acid alterations results in a catalytically inactive effector protein variant.

[186] In some embodiments, the effector proteins function as an endonuclease that catalyzes cleavage within a target nucleic acid. In some embodiments, the effector proteins are capable of catalyzing non-sequence-specific cleavage of a single stranded nucleic acid. In some embodiments, the effector proteins (*e.g.*, the effector proteins having the amino acid sequence of **TABLE 1A** and **TABLE 1B**) are activated to perform trans cleavage activity after binding of a guide nucleic acid with a target nucleic acid. This trans cleavage activity may also be referred to as "collateral" or "transcollateral" cleavage. Trans cleavage activity may be non-specific cleavage of nearby single-stranded nucleic acid by the activated effector protein, such as trans cleavage of detector nucleic acids with a detection moiety.

[187] Effector proteins disclosed herein may function as an endonuclease that catalyzes cleavage at a specific position (*e.g.*, at a specific nucleotide within a nucleic acid sequence) in a target nucleic acid. The target nucleic acid may be single stranded RNA (ssRNA), double stranded DNA (dsDNA) or single-stranded DNA (ssDNA). In some embodiments, the target nucleic acid is single-stranded DNA. In some embodiments, the target nucleic acid is single-stranded RNA. The effector proteins may provide cis cleavage activity, trans cleavage activity, nickase activity, or a combination thereof. Cis cleavage activity is cleavage of a target nucleic acid that is hybridized to a guide RNA (*e.g.*, a dual guide nucleic acid system or a sgRNA), wherein cleavage occurs within or directly adjacent to the region of the target nucleic acid that is hybridized to guide RNA. Trans cleavage activity (also referred to as transcollateral cleavage) is cleavage of ssDNA or ssRNA that is near, but not hybridized to the guide RNA. Trans cleavage may occur near, but not within or directly adjacent to, the region of the target nucleic acid that is hybridized to the guide nucleic acid. Trans cleavage activity may be triggered by the hybridization of the guide nucleic acid to the target nucleic acid. Nickase activity is a selective cleavage of one strand of a dsDNA.

Engineered Proteins

[188] In some embodiments, effector proteins disclosed herein are engineered proteins. Engineered proteins are not identical to a naturally-occurring protein. Engineered proteins may provide enhanced nuclease activity as compared to a naturally occurring nuclease. In some embodiments, a modification of the effector proteins may include addition of one or more amino acids, deletion of one or more amino acids, substitution of one or more amino acids, or combinations thereof. In some embodiments, effector

proteins disclosed herein are engineered proteins. Unless otherwise indicated, reference to effector proteins throughout the present disclosure include engineered proteins thereof.

[189] In some embodiments, effector proteins described herein can be modified with the addition of one or more heterologous peptides or heterologous polypeptides (referred to collectively herein as a heterologous polypeptide). In some embodiments, an effector protein modified with the addition of one or more heterologous peptides or heterologous polypeptides may be referred to herein as a fusion protein. Such fusion proteins are described herein and throughout.

[190] In some embodiments, heterologous polypeptides comprise at least two different polypeptide sequences that are not found similarly connected to one another in a native nucleic acid or protein. In some embodiments, a heterologous system comprises at least one component that is not naturally occurring together with remaining components of the heterologous system.

[191] In some embodiments, fusion proteins are targeted by a guide nucleic acid (guide RNA) to a specific location in the target nucleic acid. In such embodiments, a fusion partner comprises a subcellular localization signal. In some embodiments, a subcellular localization signal can be a nuclear localization signal (NLS). In some embodiments, the NLS facilitates localization of a nucleic acid, protein, or small molecule to the nucleus, when present in a cell that contains a nuclear compartment. **TABLE 2** lists exemplary NLS sequences. In some embodiments, the subcellular localization signal is a nuclear export signal (NES), a sequence to keep an effector protein retained in the cytoplasm, a mitochondrial localization signal for targeting to the mitochondria, a chloroplast localization signal for targeting to a chloroplast, an ER retention signal, and the like. In some embodiments, an effector protein described herein is not modified with a subcellular localization signal so that the polypeptide is not targeted to the nucleus, which can be advantageous depending on the circumstance (*e.g.*, when the target nucleic acid is an RNA that is present in the cytosol).

[192] In some embodiments, the heterologous polypeptide is an endosomal escape peptide (EEP). An EEP is an agent that quickly disrupts the endosome in order to minimize the amount of time that a delivered molecule, such an effector protein, spends in the endosome-like environment, and to avoid getting trapped in the endosomal vesicles and degraded in the lysosomal compartment. An exemplary EEP is set forth in **TABLE 2**.

[193] In some embodiments, the heterologous polypeptide is a cell penetrating peptide (CPP), also known as a Protein Transduction Domain (PTD). A CPP or PTD is a polypeptide, polynucleotide, carbohydrate, or organic or inorganic compound that facilitates traversing a lipid bilayer, micelle, cell membrane, organelle membrane, or vesicle membrane.

[194] Further suitable heterologous polypeptides include, but are not limited to, proteins (or fragments/domains thereof) that are boundary elements (*e.g.*, CTCF), proteins and fragments thereof that provide periphery recruitment (*e.g.*, Lamin A, Lamin B, etc.), and protein docking elements (*e.g.*, FKBP/FRB, Pil1/Aby1, etc.).

[195] In some embodiments, a heterologous peptide or heterologous polypeptide comprises a protein tag. In some embodiments, the protein tag is referred to as purification tag or a fluorescent protein. The protein tag may be detectable for use in detection of the effector protein and/or purification of the effector protein. Accordingly, in some embodiments, compositions, systems and methods comprise a protein tag or use thereof. Any suitable protein tag may be used depending on the purpose of its use. Non-limiting examples of protein tags include a fluorescent protein, a histidine tag, *e.g.*, a 6XHis tag; a hemagglutinin (HA) tag; a FLAG tag; a Myc tag; and maltose binding protein (MBP). In some embodiments, the protein tag is a portion of MBP that can be detected and/or purified. Non-limiting examples of fluorescent proteins include green fluorescent protein (GFP), yellow fluorescent protein (YFP), red fluorescent protein (RFP), cyan fluorescent protein (CFP), mCherry, and tdTomato.

[196] A heterologous polypeptide may be located at or near the amino terminus (N-terminus) of the effector protein disclosed herein. A heterologous polypeptide may be located at or near the carboxy terminus (C-terminus) of the effector proteins disclosed herein. In some embodiments, a heterologous polypeptide is located internally in an effector protein described herein (i.e., is not at the N- or C-terminus of an effector protein described herein) at a suitable insertion site.

[197] In some embodiments, heterologous peptides or heterologous polypeptides may be located at or near the amino terminus (N-terminus) of the effector protein disclosed herein. In some embodiments, heterologous peptides or heterologous polypeptides may be located at or near the carboxy terminus (C-terminus) of the effector proteins disclosed herein. In some embodiments, heterologous peptides or heterologous polypeptides are located internally in an effector protein described herein (i.e., is not at the N- or C- terminus of an effector protein described herein) at a suitable insertion site. In some embodiments, an effector protein described herein comprises 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more heterologous polypeptides at or near the N-terminus, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more heterologous polypeptides at or near the C-terminus, or a combination of these (*e.g.*, one or more heterologous polypeptides at the amino-terminus and one or more heterologous polypeptides at the carboxy terminus). When more than one heterologous polypeptide is present, each may be selected independently of the others, such that a single heterologous polypeptide may be present in more than one copy and/or in combination with one or more other heterologous polypeptides present in one or more copies. In some embodiments, a heterologous polypeptide is considered near the N- or C-terminus when the nearest amino acid of the heterologous polypeptide is within about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, or more amino acids along the polypeptide chain from the N- or C-terminus.

[198] In some embodiments, a subcellular localization signal described herein comprises any one of the amino acid sequences recited in **TABLE 2**. In some embodiments, effector proteins described herein comprise an amino acid sequence that at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 97%, or at least 98%, at least about 99%, or about 100% identical to any one of the amino acid sequences recited

in **TABLE 1A** and **TABLE 1B** and further comprises one or more amino acid sequences recited in **TABLE 2**.

[199] In some embodiments, compositions, methods and systems described herein may comprise a nuclear localization signal sequence that is connected to the N terminal of the effector protein or that is connected to the C terminal of the effector protein, or both. **TABLE 2** recites exemplary nuclear localization sequences. In some embodiments, the nuclear localization signal comprises at least one of the amino acid sequences recited in **TABLE 2.** In some embodiments, the nuclear localization signal comprises at least one of the amino acid sequences of **SEQ ID NO: 232** and **233**. In some embodiments, the nuclear localization signal comprises the amino acid sequence of **SEQ ID NO: 232**. In some embodiments, the nuclear localization signal comprises the amino acid sequence of **SEQ ID NO: 233**. In certain embodiments, the nucleotide sequence encoding the effector protein is codon optimized (*e.g.*, for expression in a eukaryotic cell) relative to the naturally occurring sequence.

[200] In certain embodiments, compositions comprise an effector protein, or a nucleic acid encoding the effector protein, wherein the effector protein comprises an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identical to any one of **SEQ ID NO: 1-231**, wherein the effector protein further comprises one or more nuclear localization signal sequence. In some embodiments, the amino acid sequence of the effector protein is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of **SEQ ID NO: 1-231,** wherein the effector protein further comprises one or more nuclear localization signal sequence. In certain embodiments, compositions comprise an effector protein, or a nucleic acid encoding the effector protein, wherein the effector protein comprises an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identical to any one of **SEQ ID NO: 1-231**, wherein at least one of the N or C terminus comprises at least one of nuclear localization signal sequence recited in **TABLE 2**. In some embodiments, the amino acid sequence of the effector protein is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of **SEQ ID NO: 1- 231**, wherein at least one of the N or C terminus comprises at least one of the nuclear localization signal sequences recited in **TABLE 2**.

[201] In some embodiments, a fusion partner is an endosomal escape peptide. An endosomal escape peptide is an agent that quickly disrupts the endosome in order to minimize the amount of time that a delivered molecule, such an effector protein, spends in the endosome-like environment, and to avoid getting trapped in the endosomal vesicles and degraded in the lysosomal compartment. In some embodiments, the fusion partner is an endosomal escape peptide. In some embodiments, an endosomal escape protein comprises the amino acid sequence GLFXALLXLLXSLWXLLLXA **(SEQ ID NO: 1822)**, wherein each X is independently selected from lysine, histidine, and arginine. In some embodiments, an endosomal escape protein comprises the amino acid sequence

GLFHALLHLLHSLWHLLLHA **(SEQ ID NO: 1823)**. In some embodiments, the amino acid sequence of the endosomal escape protein is GLFXALLXLLXSLWXLLLXA **(SEQ ID NO: 1822),** wherein each X is independently selected from lysine, histidine, and arginine or GLFHALLHLLHSLWHLLLHA **(SEQ ID NO: 1823)**.

[202] In some embodiments, effector proteins described herein are encoded by a codon optimized nucleic acid. In some embodiments, compositions, systems, and methods described herein comprise a nucleic acid encoding an effector protein described herein, wherein the nucleic acid is codon optimized. In some embodiments, a nucleic acid encoding an effector protein described herein may be codon optimized for expression in a specific cell (*e.g.*, a eukaryotic cell, an animal cell, a mammalian cell, or a human cell). In some embodiments, a nucleic acid encoding the effector protein described herein is codon optimized for a human cell. In some embodiments the nucleic acid encoding the effector protein described herein comprises a mutation to mimic the codon preferences of the intended host organism or cell while encoding the same effector protein. Thus, the codons can be changed, but the encoded effector protein remains unchanged. For example, if the intended target cell was a human cell, a human codon-optimized nucleotide sequence encoding an effector protein could be used. As another non-limiting example, if the intended host cell were a mouse cell, then a mouse codon-optimized nucleotide sequence encoding an effector protein could be generated. As another non-limiting example, if the intended host cell were a eukaryotic cell, then a eukaryote codon-optimized nucleotide sequence encoding an effector protein could be generated. As another non-limiting example, if the intended host cell were a prokaryotic cell, then a prokaryote codon-optimized nucleotide sequence encoding an effector protein could be generated. Codon usage tables are readily available, for example, at the "Codon Usage Database" available at www.kazusa.or.jp/codon.

[203] In some embodiments, compositions, methods and systems described herein may comprise a nuclear localization signal sequence that is connected to the N terminal of the effector protein or that is connected to the C terminal of the effector protein, or both. **TABLE 2** recites exemplary nuclear localization sequences. In some embodiments, the nuclear localization signal comprises at least one of the amino acid sequences recited in **TABLE 2.** In some embodiments, the nuclear localization signal comprises at least one of the amino acid sequences recited in **SEQ ID NO: 232** and **233**. In some embodiments, the nuclear localization signal comprises the amino acid sequence recited in **SEQ ID NO: 232**. In some embodiments, the nuclear localization signal comprises the amino acid sequence recited in **SEQ ID NO: 233**. In certain embodiments, the nucleotide sequence encoding the effector protein is codon optimized (*e.g.,* for expression in a eukaryotic cell) relative to the naturally occurring sequence.

[204] In some embodiments, effector proteins may comprise one or more modifications that may provide altered activity as compared to a WT counterpart (*e.g.*, a WT nuclease or nickase). In some embodiments, activity (*e.g.*, nickase, nuclease, binding, *etc.*, activity) of effector proteins described herein can be measured relative to a WT effector protein or compositions containing the same in a

cleavage assay. For example, effector proteins may comprise one or more modifications that may provide increased activity as compared to a WT counterpart. As another example, effector proteins may provide increased catalytic activity (*e.g.*, nuclease or nickase activity) as compared to a WT counterpart. Effector proteins may provide enhanced nucleic acid binding activity (*e.g.*, enhanced binding of a guide nucleic acid and/or target nucleic acid) as compared to a WT counterpart. An effector protein may have a 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 120%, 140%, 160%, 180%, 200%, or more, increase of the activity of a WT counterpart. In some embodiments, a cleavage assay is an assay designed to visualize, quantitate or identify cleavage of a nucleic acid.

Nuclease-dead effector proteins

[205] An engineered protein may comprise a modified form of a wildtype counterpart protein (*e.g.*, an effector protein). In some embodiments, the modified form of the wildtype counterpart may comprise an amino acid change (*e.g.*, deletion, insertion, substitution, or combination thereof) that reduces the nucleic acid-cleaving activity of the effector protein relative to the wildtype counterpart. For example, a nuclease domain (*e.g.*, RuvC domain) of an effector protein may be deleted or mutated relative to a wildtype counterpart effector protein so that it is no longer functional or comprises reduced nuclease activity. The modified form of the effector protein may have less than 90%, less than 80%, less than 70%, less than 60%, less than 50%, less than 40%, less than 30%, less than 20%, less than 10%, less than 5%, or less than 1% of the nucleic acid-cleaving activity of the wild-type counterpart. Engineered proteins may have no substantial nucleic acid-cleaving activity. Engineered proteins may be enzymatically inactive or "dead," that is it may bind to a nucleic acid but not cleave it. An enzymatically inactive protein may comprise an enzymatically inactive domain (*e.g.*, inactive nuclease domain). Enzymatically inactive may refer to an activity less than 1%, less than 2%, less than 3%, less than 4%, less than 5%, less than 6%, less than 7%, less than 8%, less than 9%, or less than 10% activity compared to the wild-type counterpart. A dead protein may associate with a guide nucleic acid to activate or repress transcription of a target nucleic acid. In some embodiments, the enzymatically inactive protein is fused with a protein comprising recombinase activity. In some embodiments, an amino acid change in the wildtype effector protein preserves the same nucleic acid-cleaving activity relative to the wildtype counterpart. In some embodiments, an amino acid change in the wildtype effector protein may increase nucleic acid-cleaving activity relative to the wildtype counterpart. In such embodiments, the modified form of the effector protein may have 5%, 10%, 20%, 40%, 60%, 80%, 100% or more nucleic acid-cleaving activity relative to the wildtype counterpart.

[206] An effector protein that has decreased catalytic activity relative to a naturally occurring counterpart may be referred to as catalytically inactive, enzymatically inactive, catalytically dead, or enzymatically dead. An effector protein that has decreased catalytic activity relative to a naturally occurring counterpart may be referred to as a dead protein or a dCas protein. In some embodiments, a dCas protein comprises an enzymatically inactive domain (*e.g.* inactive nuclease domain). For example,

a nuclease domain (*e.g.*, RuvC domain) of an effector protein may be deleted or mutated relative to a wildtype counterpart so that it is no longer functional or comprises reduced nuclease activity. In some embodiments, effector proteins disclosed herein comprise a truncated RuvC domain. In some embodiments, reduced nuclease activity may refer to an activity less than 10%, less than 20%, less than 30%, less than 40%, less than 50%, less than 60%, less than 70%, less than 80%, less than 90%, or less than 100% activity compared to the wild-type counterpart. In some embodiments, a catalytically inactive effector protein may bind to a guide nucleic acid and/or a target nucleic acid but does not cleave the target nucleic acid. In some embodiments, a catalytically inactive effector protein may associate with a guide nucleic acid to activate or repress transcription of a target nucleic acid. In some embodiments, a catalytically inactive effector protein is fused to an effector partner that confers an alternative activity to an effector protein activity. Such fusion proteins are described herein and throughout.

[207] In some embodiments, the effector protein may comprise an enzymatically inactive and/or "dead" (abbreviated by "d") effector protein in combination (*e.g.*, fusion) with a polypeptide comprising recombinase activity. In some embodiments, nuclease-dead effector protein may also be referred to as a catalytically inactive effector protein. Although an effector protein normally has nuclease activity, in some embodiments, an effector protein does not have nuclease activity. In some embodiments, an effector protein comprising a nuclease-dead effector protein, wherein the nuclease-dead effector protein comprising 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 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the effector protein comprising 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 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**, wherein the effector protein is modified or engineered to be a nuclease-dead effector protein.

[208] The nuclease-dead effector proteins (*e.g.*, catalytically inactive effector proteins), as described herein may comprise a modified form of a wildtype counterpart. The modified form of the wildtype counterpart may comprise an amino acid change (*e.g.*, deletion, insertion, or substitution) that reduces the nucleic acid-cleaving activity of the effector protein. In such embodiments, the catalytically inactive effector protein may also be referred to as a catalytically reduced effector protein. For example, a nuclease domain (*e.g.*, HEPN domain, RuvC domain) of an effector protein can be deleted or mutated so that it is no longer functional or comprises reduced nuclease activity. The modified form of the effector protein may have less than 90%, less than 80%, less than 70%, less than 60%, less than 50%, less than 40%, less than 30%, less than 20%, less than 10%, less than 5%, or less than 1% of the nucleic acid-cleaving activity of the wild-type counterpart. The modified form of an effector protein may have no substantial nucleic acid-cleaving activity. When an effector protein is a modified form that has no

substantial nucleic acid-cleaving activity, it may be referred to as enzymatically inactive and/or dead. A dead effector polypeptide (*e.g.*, catalytically inactive effector protein) may bind to a target nucleic acid but may not cleave the target nucleic acid. A dead effector polypeptide (*e.g.*, catalytically inactive effector protein) may associate with a guide nucleic acid to activate or repress transcription of a target nucleic acid.

[209] In some embodiments, a nuclease-dead effector protein 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 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 6**, and wherein the effector protein further comprises one or more alterations selected from D369A, D369N, E567A, E567Q, D658A and D658N. In some embodiments, a nuclease-dead effector protein 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 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% similar to **SEQ ID NO: 6**, and wherein the effector protein further comprises one or more alterations selected from D369A, D369N, E567A, E567Q, D658A and D658N.

[210] In some embodiments, a nuclease-dead effector protein 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 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 6**, and wherein the effector protein further comprises E567A substitution. In some embodiments, a nuclease-dead effector protein 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 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 6**, and wherein the effector protein further comprises E567A substitution.

[211] In some embodiments, a nuclease-dead effector protein 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 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 6**, and wherein the effector protein further comprises E567Q substitution. In some embodiments, a nuclease-dead effector protein 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 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 6**, and wherein the effector protein further comprises E567Q substitution.

[212] In some embodiments, a nuclease-dead effector protein 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 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 7**, and wherein the effector proteins comprise alanine amino acid at position 369. In some embodiments, a nuclease-dead effector protein 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 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% similar to **SEQ ID NO: 7**, and wherein the effector proteins

comprise alanine amino acid at position 369. In some embodiments, a nuclease-dead effector protein 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 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 8**, and wherein the effector protein comprises asparagine amino acid at position 369. In some embodiments, a nuclease-dead effector protein 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 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% similar to **SEQ ID NO: 8**, and wherein the effector protein comprises asparagine amino acid at position 369. In some embodiments, a nuclease-dead effector protein 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 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 9**, and wherein the effector proteins comprise alanine amino acid at position 567. In some embodiments, a nuclease-dead effector protein 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 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% similar to **SEQ ID NO: 9**, and wherein the effector proteins comprise alanine amino acid at position 567. In some embodiments, a nuclease-dead effector protein 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 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 10**, and wherein the effector protein comprises glutamine amino acid at position 567. In some embodiments, a nuclease-dead effector protein 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 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% similar to **SEQ ID NO: 10**, and wherein the effector protein comprises glutamine amino acid at position 567. In some embodiments, a nuclease-dead effector protein 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 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 11**, and wherein the effector proteins comprise alanine amino acid at position 658. In some embodiments, a nuclease-dead effector protein 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 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% similar to **SEQ ID NO: 11**, and wherein the effector proteins comprise alanine amino acid at position 658. In some embodiments, a nuclease-dead effector protein 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 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 12**, and wherein the effector protein comprises asparagine amino acid at position 658. In some embodiments, a nuclease-dead effector protein 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 92%, at least

95%, at least 97%, at least 98%, at least 99%, or 100% similar to **SEQ ID NO: 12**, and wherein the effector protein comprises asparagine amino acid at position 658.

[213] In some embodiments, a nuclease-dead effector protein 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 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 231**, and wherein the effector protein further comprises one or more alterations selected from D237A, D418A, D418N, E335A, and E335Q. In some embodiments, a nuclease-dead effector protein 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 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% similar to **SEQ ID NO: 231**, and wherein the effector protein further comprises one or more alterations selected from D237A, D418A, D418N, E335A, and E335Q.

[214] In some embodiments, a nuclease-dead effector protein 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 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 231**, and wherein the effector protein further comprises D237A substitution. In some embodiments, a nuclease-dead effector protein 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 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% similar to **SEQ ID NO: 231**, and wherein the effector protein further comprises D237A substitution.

[215] In some embodiments, a nuclease-dead effector protein 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 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 231**, and wherein the effector protein further comprises E335Q substitution. In some embodiments, a nuclease-dead effector protein 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 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% similar to **SEQ ID NO: 231**, and wherein the effector protein further comprises E335Q substitution.

Fusion Proteins

[216] In some embodiments, compositions, systems, and methods comprise a fusion protein or uses thereof. A fusion protein generally comprises an effector protein and a fusion partner protein. A fusion partner protein is also simply referred to herein as a fusion partner. In some embodiments, the fusion partner is a heterologous protein capable of imparting some function or activity that is not provided by the effector protein. In some embodiments, the fusion partner comprises a polypeptide or peptide that is fused or linked to the effector protein. In some embodiments, the fusion partner is fused or linked to the effector protein. In some embodiments, the fusion partner is fused to the N-terminus of the effector protein. In some embodiments, the fusion partner is fused to the C-terminus of the effector protein. In

some embodiments, the fusion protein is a heterologous peptide or polypeptide as described herein. In some embodiments, the amino terminus of the fusion partner is linked/fused to the carboxy terminus of the effector protein. In some embodiments, the carboxy terminus of the fusion partner protein is linked/fused to the amino terminus of the effector protein by the linker. In some embodiments, the fusion partner is not an effector protein as described herein. In some embodiments, the fusion partner comprises an effector protein as described herein or a multimeric form thereof. Accordingly, in some embodiments, the fusion protein comprises more than one effector protein. In such embodiments, the fusion protein comprises at least two effector proteins that are same. In some embodiments, the fusion protein comprises at least two effector proteins that are different. In some embodiments, the multimeric form is a homomeric form. In some embodiments, the multimeric form is a heteromeric form. Unless otherwise indicated, reference to effector proteins throughout the present disclosure include fusion proteins comprising the effector protein described herein and a fusion partner.

[217] In some embodiments, a fusion partner may inhibit or promote the formation of multimeric complex of an effector protein. In an additional example, the fusion partner may directly or indirectly edit a target nucleic acid. Modifications can be of a nucleobase, nucleotide, or nucleotide sequence of a target nucleic acid. In some embodiments, the fusion partner may interact with additional proteins, or functional fragments thereof, to make modifications to a target nucleic acid.

[218] In some embodiments, an effector protein is a fusion protein, wherein the fusion protein comprises an effector protein and a fusion partner protein. In some embodiments, the effector protein, or a nucleic acid encoding the effector protein, wherein the effector protein comprises an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the amino acid of the effector protein is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. Unless otherwise indicated, reference to effector proteins throughout the present disclosure include fusion proteins thereof.

[219] In some embodiments, the fusion proteins are multimeric proteins. In some embodiments, the multimeric protein is a homomeric protein. In some embodiments, the multimeric protein is a heteromeric protein. In some embodiments, the fusion protein can comprise at least two effector proteins that are same. In some embodiments, the fusion protein comprises at least two effector proteins that are different. Unless otherwise indicated, reference to effector proteins throughout the present disclosure include fusion proteins described herein. In some embodiments, the fusion partner promotes the formation of a multimeric complex of the effector protein. In some embodiments, a fusion partner may inhibit or promote the formation of multimeric complex of an effector protein. In an additional example, the fusion partner may directly or indirectly edit a target nucleic acid. Modifications can be of a nucleobase, nucleotide, or nucleotide sequence of a target nucleic acid. In some embodiments, the fusion partner may interact with additional proteins, or functional fragments thereof, to make

modifications to a target nucleic acid.In some embodiments, the fusion protein complexes with a guide nucleic acid and the complex interacts with the target nucleic acid. In some embodiments, the interaction comprises one or more of: recognition of a PAM sequence within the target nucleic acid by the effector protein, hybridization of the guide nucleic acid to the target nucleic acid, modification of the target nucleic acid by the fusion protein, or combinations thereof. In some embodiments, recognition of a PAM sequence within a target nucleic acid may direct the modification activity of a fusion protein.

[220] Modification activity of a fusion protein described herein may be cleavage activity, binding activity, insertion activity, substitution activity, and the like. Modification activity of an effector protein may result in: cleavage of at least one strand of a target nucleic acid, deletion of one or more nucleotides of a target nucleic acid, insertion of one or more nucleotides into a target nucleic acid, substitution of one or more nucleotides of a target nucleic acid with an alternative nucleotide, more than one of the foregoing, or any combination thereof. In some embodiments, an ability of a fusion protein to edit a target nucleic acid may depend upon the effector protein being complexed with a guide nucleic acid, the guide nucleic acid being hybridized to a target sequence of the target nucleic acid, the distance between the target sequence and a PAM sequence, or combinations thereof.

[221] In some embodiments, the fusion protein described herein comprises a heterologous amino acid sequence that affects formation of a multimeric complex of the fusion protein. By way of non-limiting example, the fusion protein may comprise an effector protein described herein and a fusion partner comprising a Calcineurin A tag, wherein the fusion protein dimerizes in the presence of Tacrolimus (FK506). Also, by way of non-limiting example, the fusion protein may comprise an effector protein described herein and a SpyTag configured to dimerize or associate with another effector protein in a multimeric complex. Multimeric complex formation is further described herein.

[222] In some embodiments, fusion proteins are useful for precision editing of the genome. In some embodiments, the fusion partner is a reverse transcriptase and the system/composition comprises a guide nucleic acid comprising a template sequence to be incorporated into a genome. In some embodiments, the fusion partner is a base editing enzyme. In general, a base editor comprises a deaminase fused to an effector protein. In some embodiments, the base editor changes a nucleobase to a different nucleobase, *e.g.*, cytosine to thymine or guanine to adenine. In some embodiments, the base editor comprises a dCas protein. In some embodiments, the base editor comprises a Cas nickase.

Nucleic Acid Editing Activity

[223] In some embodiments, fusion partners provide enzymatic activity that modifies a target nucleic acid. Such enzymatic activities include, but are not limited to, nuclease activity, demethylase activity, DNA repair activity, deamination activity, dismutase activity, integrase activity, transposase activity, recombinase activity, polymerase activity, ligase activity, helicase activity, photolyase activity, and glycosylase activity.

[224] In some embodiments, fusion partners have enzymatic activity that modifies the target nucleic acid. The target nucleic acid may comprise or consist of a ssRNA, dsRNA, ssDNA, or a dsDNA. Examples of enzymatic activity that modifies the target nucleic acid include, but are not limited to: nuclease activity, which comprises the enzymatic activity of an enzyme which allows the enzyme to cleave the phosphodiester bonds between the nucleotide subunits of nucleic acids, such as that provided by a restriction enzyme (*e.g.*, FokI nuclease); demethylase activity such as that provided by a demethylase (*e.g.*, Ten-Eleven Translocation (TET) dioxygenase 1 (TET1CD), TET1, DME, DML1, DML2, ROS1); DNA repair activity; deamination activity such as that provided by a deaminase (*e.g.*, a cytosine deaminase enzyme such as rat APOBEC1); dismutase activity; integrase activity such as that provided by an integrase and/or resolvase (*e.g.*, Gin invertase such as the hyperactive mutant of the Gin invertase, GinH106Y; human immunodeficiency virus type 1 integrase (IN); Tn3 resolvase); transposase activity, recombinase activity such as that provided by a recombinase (*e.g.*, catalytic domain of Gin recombinase); as well as polymerase activity, ligase activity, helicase activity, photolyase activity, and glycosylase activity. In some embodiments, transposase activity refers to catalytic activity that results in the transposition of a first nucleic acid into a second nucleic acid.

[225] Non-limiting examples of fusion partners for targeting ssRNA include, but are not limited to, protein translation components (*e.g.*, translation initiation, elongation, and/or release factors; *e.g.*, eIF4G); helicases; and RNA-binding proteins. It is understood that a fusion protein may include the entire protein or in some embodiments may include a fragment of the protein (*e.g.*, a functional domain). In some embodiments, the functional domain interacts with or binds ssRNA, including intramolecular and/or intermolecular secondary structures thereof, *e.g.*, hairpins, stem-loops, etc.). The functional domain may interact transiently or irreversibly, directly or indirectly.

[226] In some embodiments, a functional domain comprises a region of one or more amino acids in a protein that is required for an activity of the protein, or the full extent of that activity. Activities include but are not limited to nucleic acid binding, nucleic acid editing, nucleic acid modifying, nucleic acid cleaving, protein binding or combinations thereof. The absence of the functional domain, including mutations of the functional domain, would abolish or reduce activity.

[227] Fusion proteins may comprise a protein or domain thereof selected from: endonucleases (*e.g.*, RNase III, the CRR22 DYW domain, Dicer, and PIN (PilT N-terminus); SMG5 and SMG6; exonucleases such as XRN-1 or Exonuclease T; proteins and protein domains responsible for RNA localization (*e.g.*, from IMP1, ZBP1, She2p, She3p, and Bicaudal-D); proteins and protein domains responsible for nuclear retention of RNA (*e.g.*, Rrp6); proteins and protein domains responsible for nuclear export of RNA (*e.g.*, TAP, NXF1, THO, TREX, REF, and Aly). Alternatively, the effector domain may be a domain of a protein selected from the group comprising endonucleases; proteins and protein domains capable of stimulating RNA cleavage; exonucleases; and proteins and protein domains having RNA localization activity; proteins and protein domains capable of nuclear retention of RNA; proteins and protein domains having RNA nuclear export activity.

Recombinases

[228] In some embodiments, fusion partners comprise a recombinase. In some embodiments, effector proteins described herein are fused with the recombinase. In some embodiments, the effector proteins have reduced nuclease activity or no nuclease activity. In some embodiments, the recombinase is a site-specific recombinase.

[229] In some embodiments, an enzymatically inactive effector protein is fused with a recombinase, wherein the recombinase can be a site-specific recombinase. Such polypeptides can be used for site-directed transgene insertion. Non-limiting examples of site-specific recombinases include a tyrosine recombinase (*e.g.*, Cre, Flp or lambda integrase), a serine recombinase (*e.g.*, gamma-delta resolvase, Tn3 resolvase, Sin resolvase, Gin invertase, Hin invertase, Tn5044 resolvase, IS607 transposase and integrase), or mutants or variants thereof. In some embodiments, the recombinase is a serine recombinase. Non-limiting examples of serine recombinases include: gamma-delta resolvase, Tn3 resolvase, Sin resolvase, Gin invertase, Hin invertase, Tn5044 resolvase, IS607 transposase, and IS607 integrase. In some embodiments, the site-specific recombinase is an integrase. Non-limiting examples of integrases include: Bxb1, wBeta, BL3, phiR4, A118, TG1, MR11, phi370, SPBc, TP901-1, phiRV, FC1, K38, phiBT1, and phiC31. Further discussion and examples of suitable recombinase fusion partners are described in US 10,975,392, which is incorporated herein by reference in its entirety. In some embodiments, the fusion protein comprises a linker that links the recombinase to the Cas-CRISPR domain of the effector protein. In some embodiments, the linker is The-Ser.

Polypeptide Linkers

[230] In some embodiments, a linker comprises a bond or molecule that links a first polypeptide to a second polypeptide. In some embodiments, effector proteins, fusion proteins, fusion partners, or combinations thereof are connected by linkers. In some embodiments, effector proteins and fusion partners of a fusion effector protein are connected by a linker. The linker may comprise or consist of a covalent bond. The linker may comprise or consist of a chemical group. In some embodiments, the linker comprises an amino acid. In some embodiments, a peptide linker comprises at least two amino acids linked by an amide bond. In general, the linker connects a terminus of the effector protein to a terminus of the fusion partner. In some embodiments, carboxy terminus of the effector protein is linked to the amino terminus of the fusion partner. In some embodiments, carboxy terminus of the fusion partner is linked to the amino terminus of the effector protein. In some embodiments, the effector protein and the fusion partner are directly linked by a covalent bond.

[231] In some embodiments, linkers comprise one or more amino acids. In some embodiments, a linker comprises a peptide. In some embodiments, a terminus of the effector protein is linked to a terminus of the fusion partner through an amide bond. In some embodiments, a terminus of the effector protein is linked to a terminus of the fusion partner through a peptide bond. In some embodiments, linkers comprise an amino acid. In some embodiments, linkers comprise one or more amino acids. In

some embodiments, linkers comprise a peptide. In some embodiments, linker is a protein. In some embodiments, an effector protein is coupled to a fusion partner via a linker protein. The linker protein may have any of a variety of amino acid sequences. A linker protein may comprise a region of rigidity (*e.g.*, beta sheet, alpha helix), a region of flexibility, or any combination thereof. In some embodiments, the linker comprises small amino acids, such as glycine and alanine, that impart high degrees of flexibility. The ordinarily skilled artisan will recognize that design of a peptide conjugated to any desired element may include linkers that are all or partially flexible, such that the linker may include a flexible linker as well as one or more portions that confer less flexible structure. Suitable linkers include proteins of 4 linked amino acids to 40 linked amino acids in length, or between 4 linked amino acids and 25 linked amino acids in length. In some embodiments, linked amino acids is described herein comprise at least two amino acids linked by an amide bond.

[232] Linkers may be produced by using synthetic, linker-encoding oligonucleotides to couple the proteins, or may be encoded by a nucleic acid sequence encoding a fusion protein (*e.g.,* an effector protein coupled to a fusion partner). In some embodiments, the linker is from 1 to 100 amino acids in length. In some embodiments, the linker is more 100 amino acids in length. In some embodiments, the linker is from 10 to 27 amino acids in length. Examples of linker proteins include glycine polymers (G)n, glycine-serine polymers (including, for example, (GS)n, GSGGSn (**SEQ ID NO:1851**), GGSGGSn (**SEQ ID NO:1852**), and GGGSn (**SEQ ID NO:1853**), where n is an integer of at least one), glycine-alanine polymers, and alanine-serine polymers. Exemplary linkers may comprise amino acid sequences including, but not limited to, GS, GSGGS (**SEQ ID NO:1851**), GGSGGS (**SEQ ID NO:1852**), GGGS (**SEQ ID NO:1853**), GGSG (**SEQ ID NO:1854**), GGSGG (**SEQ ID NO:1855**), GSGSG (**SEQ ID NO:1856**), GSGGG (**SEQ ID NO:1857**), GGGSG (**SEQ ID NO:1858**), and GSSSG (**SEQ ID NO:1859**). In some embodiments, the linker comprises one or more repeats a tri-peptide GGS. In some embodiments, the linker is an XTEN linker. In some embodiments, the linker is an XTEN linker. In some embodiments, the XTEN linker is an XTEN80 linker. In some embodiments, the XTEN linker is an XTEN20 linker. In some embodiments, the XTEN20 linker has an amino acid sequence of GSGGSPAGSPTSTEEGTSESATPGSG (**SEQ ID NO: 1790**).

[233] In some embodiments, linkers do not comprise an amino acid. In some embodiments, linkers do not comprise a peptide. In some embodiments, linkers comprise a nucleotide, a polynucleotide, a polymer, or a lipid. In some embodiments, linker may be a polyethylene glycol (PEG), polypropylene glycol (PPG), co-poly(ethylene/propylene) glycol, polyoxyethylene (POE), polyurethane, polyphosphazene, polysaccharides, dextran, polyvinyl alcohol, polyvinylpyrrolidones, polyvinyl ethyl ether, polyacrylamide, polyacrylate, polycyanoacrylates, lipid polymers, chitins, hyaluronic acid, heparin, or an alkyl linker.

[234] In some embodiments, a linker is recognized and cleaved by a protein. In some embodiments, a linker comprises a recognition sequence that may be recognized and cleaved by the protein. In some embodiments, a guide nucleic acid comprises an aptamer, which may serve a similar function as a

linker, bringing an effector protein and an effector partner protein into proximity. The aptamer can functionally connect two proteins (*e.g.,* effector protein, effector partner) by interacting non-covalently with both, thereby bringing both proteins into proximity of the guide nucleic acid. In some embodiments, the first protein and/or the second protein comprise or is covalently linked to an aptamer binding moiety. In some embodiments, the aptamer is a short single stranded DNA (ssDNA) or RNA (ssRNA) molecule capable of being bound be the aptamer binding moiety. In some embodiments, the aptamer is a molecule that is capable of mimicking antibody binding activity and may be classified as a chemical antibody. In some embodiments, the aptamer described herein refers to artificial oligonucleotides that bind one or more specific molecules. In some embodiments, aptamers exhibit a range of affinities (K_D in the pM to μM range) with little or no off-target binding.

Multimeric Complexes

[235] Compositions, systems, and methods of the present disclosure may comprise a multimeric complex or uses thereof, wherein the multimeric complex comprises multiple effector proteins that non-covalently interact with one another. A multimeric complex may comprise enhanced activity relative to the activity of any one of its effector proteins alone. For example, a multimeric complex comprising two effector proteins may comprise greater nucleic acid binding affinity, cis-cleavage activity, and/or transcollateral cleavage activity than that of either of the effector proteins provided in monomeric form. A multimeric complex may have an affinity for a target region of a target nucleic acid and is capable of catalytic activity (*e.g.*, cleaving, nicking, editing or modifying the nucleic acid) at or near the target region. Multimeric complexes may be activated when complexed with a guide nucleic acid. Multimeric complexes may be activated when complexed with a guide nucleic acid and a target nucleic acid. In some embodiments, the multimeric complex cleaves the target nucleic acid. In some embodiments, the multimeric complex nicks the target nucleic acid.

[236] Various aspects of the present disclosure include compositions and methods comprising multiple effector proteins, and uses thereof, respectively. An effector protein comprising an amino acid sequence that is at least 70% identical to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**, wherein the effector protein may be provided with a second effector protein. Two effector proteins may target different nucleic acid sequences. Two effector proteins may target different types of nucleic acids (*e.g.*, a first effector protein may target double- and single-stranded nucleic acids, and a second effector protein may only target single-stranded nucleic acids).

[237] In some embodiments, multimeric complexes comprise at least one effector protein, or a fusion protein thereof, comprising an amino acid sequence with at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% or 100% identity to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, multimeric complexes comprise at least one effector protein or a fusion protein thereof, wherein the amino acid sequence of the effector protein is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%,

at least 95% or 100% identical to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**.

[238] In some embodiments, the multimeric complex is a dimer comprising two effector proteins of identical amino acid sequences. In some embodiments, the multimeric complex comprises a first effector protein and a second effector protein, wherein the amino acid sequence of the first effector protein is at least 90%, at least 92%, at least 94%, at least 96%, at least 98% identical, or at least 99% identical to the amino acid sequence of the second effector protein.

[239] In some embodiments, the multimeric complex is a heterodimeric complex comprising at least two effector proteins of different amino acid sequences. In some embodiments, the multimeric complex is a heterodimeric complex comprising a first effector protein and a second effector protein, wherein the amino acid sequence of the first effector protein is less than 90%, less than 85%, less than 80%, less than 75%, less than 70%, less than 65%, less than 60%, less than 55%, less than 50%, less than 45%, less than 40%, less than 35%, less than 30%, less than 25%, less than 20%, less than 15%, or less than 10% identical to the amino acid sequence of the second effector protein.

[240] In some embodiments, a multimeric complex comprises at least two effector proteins. In some embodiments, a multimeric complex comprises more than two effector proteins. In some embodiments, a multimeric complex comprises two, three or four effector proteins. In some embodiments, at least one effector protein of the multimeric complex comprises an amino acid sequence with at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% or 100% identity to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, each effector protein of the multimeric complex comprises an amino acid sequence with at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% or 100% identity to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**.

Synthesis, Isolation and Assaying

[241] Effector proteins of the present disclosure may be synthesized, using any suitable method. In some embodiments, the effector proteins may be produced *in vitro* or by eukaryotic cells or by prokaryotic cells. In some embodiments, the effector proteins may be further processed by unfolding (*e.g.*, heat denaturation, dithiothreitol reduction, *etc.*) and may be further refolded, using any suitable method. In some embodiments, *in vitro* describes something outside an organism. An *in vitro* system, composition, or method may take place outside an organism and/or in a container.

[242] Any suitable method of generating and assaying effector proteins described herein may be used. Such methods include, but are not limited to, site-directed mutagenesis, random mutagenesis, combinatorial libraries, and other mutagenesis methods described herein (see, *e.g.*, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, Third Ed., Cold Spring Harbor Laboratory, New York (2001); Ausubel *et al.*, Current Protocols in Molecular Biology, John Wiley and Sons, Baltimore, MD (1999); Gillman *et al.*, Directed Evolution Library Creation: Methods and Protocols (Methods in

Molecular Biology) Springer, 2nd ed (2014)). One non-limiting example of a method for preparing an effector protein is to express recombinant nucleic acids encoding the effector protein in a suitable microbial organism, such as a bacterial cell, a yeast cell, or other suitable cell, using methods well known in the art. Exemplary methods are also described in the Examples provided herein.

[243] In some embodiments, an effector protein provided herein is an isolated effector protein. In some embodiments, the effector proteins may be isolated and purified for use in compositions, systems, and/or methods described herein. In some embodiments, methods described here may include the step of isolating effector proteins described herein. Any suitable method to provide isolated effector proteins described herein may be used in the present disclosure, for example, recombinant expression systems, precipitation, gel filtration, ion-exchange, reverse-phase and affinity chromatography, and the like. Other well-known methods are described in Deutscher *et al.*, Guide to Protein Purification: Methods in Enzymology, Vol. 182, (Academic Press, (1990)). Alternatively, the isolated polypeptides of the present disclosure can be obtained using well-known recombinant methods (see, *e.g.*, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, Third Ed., Cold Spring Harbor Laboratory, New York (2001); and Ausubel *et al.*, Current Protocols in Molecular Biology, John Wiley and Sons, Baltimore, MD (1999)). The methods and conditions for biochemical purification of a polypeptide described herein can be chosen by those skilled in the art, and purification monitored, for example, by a functional assay.

[244] In some embodiments, compositions, systems, and methods described herein may further comprise a purification tag that can be attached to an effector protein, or a nucleic acid encoding the purification tag that can be attached to a nucleic acid encoding the effector protein as described herein. In some embodiments, the purification tag may be an amino acid sequence which can attach or bind with high affinity to a separation substrate and assist in isolating the protein of interest from its environment, which may be its biological source, such as a cell lysate. Attachment of the purification tag may be at the N or C terminus of the effector protein. Furthermore, an amino acid sequence recognized by a protease or a nucleic acid encoding for an amino acid sequence recognized by a protease, such as TEV protease or the HRV3C protease may be inserted between the purification tag and the effector protein, such that biochemical cleavage of the sequence with the protease after initial purification liberates the purification tag. Purification and/or isolation may be performed through high performance liquid chromatography (HPLC), exclusion chromatography, gel electrophoresis, affinity chromatography, or other purification technique. Non-limiting examples of purification tags are as described herein.

[245] In some embodiments, effector proteins described herein are isolated from cell lysate. In some embodiments, the compositions described herein may comprise 20% or more by weight, 75% or more by weight, 95% or more by weight, or 99.5% or more by weight of an effector protein, related to the method of preparation of compositions described herein and its purification thereof, wherein percentages may be upon total protein content in relation to contaminants. Thus, in some embodiments,

the effector protein is at least 80% pure, at least 85% pure, at least 90% pure, at least 95% pure, at least 98% pure, or at least 99% pure (*e.g.*, free of contaminants, non-engineered polypeptide proteins or other macromolecules, *etc.*).

Protospacer Adjacent Motif (PAM) Sequences

[246] Effector proteins of the present disclosure may cleave or nick a target nucleic acid within or near a protospacer adjacent motif (PAM) sequence of the target nucleic acid. In some embodiments, the target nucleic acid is a double stranded nucleic acid comprising a target strand and a non-target strand. In some embodiments, cleavage occurs within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, or 50 nucleotides of a 5' or 3' terminus of a PAM sequence. In some embodiments, effector proteins described herein recognize a PAM sequence. In some embodiments, recognizing a PAM sequence comprises interacting with a sequence adjacent to the PAM. In some embodiments, a target nucleic acid comprises a target sequence that is adjacent to a PAM sequence. In some embodiments, the effector protein does not require a PAM to bind and/or cleave a target nucleic acid.

[247] In some embodiments, a target nucleic acid is a single stranded target nucleic acid comprising a target sequence. Accordingly, in some embodiments, the single stranded target nucleic acid comprises a PAM sequence described herein that is adjacent (*e.g.*, within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, or 50 nucleotides) or directly adjacent to the target sequence. In some embodiments, an RNP cleaves the single stranded target nucleic acid.

[248] In some embodiments, a target nucleic acid is a double stranded nucleic acid comprising a target strand and a non-target strand, wherein the target strand comprises a target sequence. In some embodiments, the PAM sequence is located on the target strand. In some embodiments, the PAM sequence is located on the non-target strand. In some embodiments, the PAM sequence described herein is adjacent (*e.g.*, within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, or 50 nucleotides) to the target sequence on the target strand or the non-target strand. In some embodiments, such a PAM described herein is directly adjacent to the target sequence on the target strand or the non-target strand. In some embodiments, an RNP cleaves the target strand or the non-target strand. In some embodiments, the RNP cleaves both, the target strand and the non-target strand. In some embodiments, an RNP recognizes the PAM sequence, and hybridizes to a target sequence of the target nucleic acid. In some embodiments, the RNP cleaves the target nucleic acid, wherein the RNP has recognized the PAM sequence and is hybridized to the target sequence.

[249] In some embodiments, an effector protein described herein, or a multimeric complex thereof, recognizes a PAM on a target nucleic acid. In some embodiments, multiple effector proteins of the multimeric complex recognize a PAM on a target nucleic acid. In some embodiments, at least two of the multiple effector proteins recognize the same PAM sequence. In some embodiments, at least two of the multiple effector proteins recognize different PAM sequences. In some embodiments, only one effector protein of the multimeric complex recognizes a PAM on a target nucleic acid.

[250] Effector proteins of the present disclosure, dimers thereof, and multimeric complexes thereof may cleave or nick a target nucleic acid within or near a protospacer adjacent motif (PAM) sequence of the target nucleic acid. In some embodiments, cleavage occurs within 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 nucleosides of a 5' or 3' terminus of a PAM sequence. A target nucleic acid may comprise a PAM sequence adjacent to a nucleotide sequence that is complementary to a guide nucleic acid spacer region. In some embodiments, the effector protein recognizes a PAM sequence. In some embodiments, the PAM sequence comprises at least three nucleotides. In some embodiments, the PAM sequence comprises at least four nucleotides. In some embodiments, the PAM sequence comprises at least two consecutive thymine bases. In some embodiments, the PAM sequence comprises at least three consecutive thymine bases. Each, nucleotide and nucleoside, describes sugar and base of the residue contained in the nucleic acid molecule.

[251] In some embodiments, the PAM sequence comprises any of the nucleotide sequences recited in **TABLE 3**. In some embodiments, a composition comprising an effector protein recognizes a PAM sequence comprising any of the nucleotide sequences recited in **TABLE 3**.

[252] In some embodiments, effector proteins described herein recognize any one of PAM sequences recited in **TABLE 3**. In some embodiments, effector proteins described herein recognize any one of PAM sequences of **SEQ ID NO: 234-251**. In some embodiments, the PAM sequence is within 20, 18, 16, 14, 12, 10, 8, 6, 4 or 2 bases of the target sequence. In some embodiments, compositions and systems described herein comprises an effector protein having an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identical to any one of **SEQ ID NO: 1-231**, and that recognizes any one of PAM sequences of **SEQ ID NO: 234-251**.

[253] In some embodiments, compositions and systems described herein comprises an effector protein having an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identical to any one of **SEQ ID NO: 6-12,** and **SEQ ID NO: 228-230**, and that recognizes any one of PAM sequences of **SEQ ID NO: 234-237**.

[254] In some embodiments, compositions and systems described herein comprises an effector protein having an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 13**, and that recognizes any one of PAM sequences of **SEQ ID NO: 238-247**.

[255] In some embodiments, compositions and systems described herein comprises an effector protein having an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 47**, and that recognizes any one of PAM sequences of **SEQ ID NO: 241-244**.

[256] In some embodiments, compositions and systems described herein comprises an effector protein having an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at

least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 231**, and that recognizes any one of PAM sequences of **SEQ ID NO: 245-248**.

[257] In some embodiments, compositions and systems described herein comprises an effector protein having an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identical to any one of **SEQ ID NO: 181-184**, and that recognizes any one of PAM sequences of **SEQ ID NO: 249,** and **1791-1794**.

[258] In some embodiments, compositions and systems described herein comprises an effector protein having an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 187** or **188**, and that recognizes a PAM sequence of **SEQ ID NO: 250**.

[259] In some embodiments, compositions and systems described herein comprises an effector protein having an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 185**, and that recognizes a PAM sequence of **SEQ ID NO: 251**.

II. Guide Nucleic Acids

[260] The compositions, systems, and methods of the present disclosure may comprise a guide nucleic acid or a use thereof. In general, a guide nucleic acid is a nucleic acid molecule that binds to an effector protein, thereby forming a ribonucleoprotein complex (RNP). Guide nucleic acids, when complexed with an effector protein, may bring the effector protein into proximity of a target nucleic acid. Sufficient conditions for hybridization of a guide nucleic acid to a target nucleic acid and/or for binding of a guide nucleic acid to an effector protein include in vivo physiological conditions of a desired cell type or in vitro conditions sufficient for assaying catalytic activity of a protein, polypeptide or peptide described herein, such as the nuclease activity of an effector protein. In general, a guide nucleic acid comprises a first region that is not complementary to a target nucleic acid (FR) and a second region is complementary to the target nucleic acid (SR). In some embodiments, FR is located 5' to SR (FR-SR). In some embodiments, SR is located 5' to FR (SR-FR). In some embodiments, the FR comprises one or more repeat sequences, handle sequences, intermediary sequences, linkers or combinations thereof. In some embodiments, an effector protein binds to at least a portion of the FR. In some embodiments, the SR comprises a spacer sequence, wherein the spacer sequence can interact in a sequence-specific manner with (*e.g.*, has complementarity with, or can hybridize to a target sequence in) a target nucleic acid. Guide nucleic acids may comprise DNA, RNA, or a combination thereof (*e.g.*, RNA with a thymine base). However, no matter the form the sequence is described, it is readily understood that such nucleotide sequences can be revised to be RNA or DNA, as needed, for describing a sequence within a guide nucleic acid itself or the sequence that encodes a guide nucleic acid. Similarly, disclosure of the nucleotide sequences described herein also discloses a complementary nucleotide sequence, a reverse nucleotide sequence, and the reverse complement nucleotide sequence, any one of

which can be a nucleotide sequence for use in a guide nucleic acid. In some embodiments, a guide nucleic acid sequence(s) comprises one or more nucleotide alterations at one or more positions in any one of the sequences described herein. Alternative nucleotides can be any one or more of A, C, G, T or U, or a deletion, or an insertion. Guide nucleic acids may include a chemically modified nucleobase or phosphate backbone. Guide nucleic acids may be referred to herein as a guide RNA (gRNA). However, a guide RNA is not limited to ribonucleotides, but may comprise deoxyribonucleotides and other chemically modified nucleotides.

[261] The compositions, systems, and methods of the present disclosure may comprise a guide nucleic acid, a nucleic acid encoding the guide nucleic acid, or a use thereof. Unless otherwise indicated, compositions, systems and methods comprising guide nucleic acids or uses thereof, as described herein and throughout, include DNA molecules, such as expression vectors, that encode a guide nucleic acid. Guide nucleic acids are also referred to herein as "guide RNA." A guide nucleic acid, as well as any components thereof (*e.g.*, spacer sequence, repeat sequence, linker nucleotide sequence, handle sequence, intermediary sequence etc.) may comprise one or more deoxyribonucleotides (DNA), ribonucleotides (RNA), a combination thereof (*e.g.*, RNA with a thymine base), biochemically or chemically modified nucleotides (*e.g.*, one or more engineered modifications as described herein), or any combinations thereof. Such nucleotide sequences described herein may be described as a nucleotide sequence of either DNA or RNA, however, no matter the form the sequence is described, it is readily understood that such nucleotide sequences can be revised to be RNA or DNA, as needed, for describing a sequence within a guide nucleic acid itself or the sequence that encodes a guide nucleic acid, such as a nucleotide sequence described herein for a vector. The exchange of uridine for thymidine or vice versa, and the presence of nucleoside analogs, such as modified uridines, do not contribute to differences in identity or complementarity among polynucleotides as long as the relevant nucleotides (such as thymidine, uridine, or modified uridine) have the same complement (*e.g.*, adenosine for all of thymidine, uridine, or modified uridine; another example is cytosine and 5- methylcytosine, both of which have guanosine or modified guanosine as a complement). Thus, for example, the sequence 5'-AXG where X is any modified uridine, such as pseudouridine, NI-methyl pseudouridine, or 5-methoxyuridine, is considered 100% identical to AUG in that both are perfectly complementary to the same sequence (5' -CAU). Similarly, disclosure of the nucleotide sequences described herein also discloses the complementary nucleotide sequence, the reverse nucleotide sequence, and the reverse complement nucleotide sequence, any one of which can be a nucleotide sequence for use in a guide nucleic acid as described herein.

[262] A guide nucleic acid may comprise a naturally occurring guide nucleic acid. A guide nucleic acid may comprise a non-naturally occurring guide nucleic acid, including a guide nucleic acid that is designed to contain a chemical or biochemical modification. A guide nucleic acid of the present disclosure comprises one or more of the following: a) a single nucleic acid molecule; b) a DNA base; c) an RNA base; d) a modified base; e) a modified sugar; f) a modified backbone; and the like. The

guide RNA may be chemically synthesized or recombinantly produced. The sequence of the guide nucleic acid, or a portion thereof, may be different from the sequence of a naturally occurring nucleic acid. Guide nucleic acids and portions thereof may be found in or identified from a CRISPR array present in the genome of a host organism.

[263] The guide nucleic acid may also form complexes as described through herein. A guide nucleic acid refers to a nucleic acid that, when in a complex with one or more polypeptides described herein (*e.g.,* an RNP complex) can impart sequence selectivity to the complex when the complex interacts with a target nucleic acid. For example, a guide nucleic acid may hybridize to another nucleic acid, such as target nucleic acid, or a portion thereof. In another example, a guide nucleic acid may complex with an effector protein. In such embodiments, a guide nucleic acid-effector protein complex may be described herein as an RNP. In some embodiments, when in a complex, at least a portion of the complex may bind, recognize, and/or hybridize to a target nucleic acid. For example, when a guide nucleic acid and an effector protein are complexed to form an RNP, at least a portion of the guide nucleic acid hybridizes to a target sequence in a target nucleic acid. Those skilled in the art in reading the below specific examples of guide nucleic acids as used in RNPs described herein, will understand that in some embodiments, a RNP may hybridize to one or more target sequences in a target nucleic acid, thereby allowing the RNP to modify and/or recognize a target nucleic acid or sequence contained therein (*e.g.*, PAM) or to modify and/or recognize non-target sequences depending on the guide nucleic acid, and in some embodiments, the effector protein, used.

[264] In some embodiments, a guide nucleic acid may comprise or form intramolecular secondary structure (*e.g.*, hairpins, stem-loops, *etc.*). In some embodiments, a guide nucleic acid comprises a stem-loop structure comprising a stem region and a loop region. In some embodiments, the stem region is 4 to 8 linked nucleotides in length. In some embodiments, the stem region is 5 to 6 linked nucleotides in length. In some embodiments, the stem region is 4 to 5 linked nucleotides in length. In some embodiments, the guide nucleic acid comprises a pseudoknot (*e.g.*, a secondary structure comprising a stem, at least partially, hybridized to a second stem or half-stem secondary structure). A nucleic acid having multiple linked nucleotides or nucleosides describes linked sugars and bases of residues contained in the nucleic acid molecule.

[265] An effector protein may recognize a guide nucleic acid comprising multiple stem regions. In some embodiments, the nucleotide sequences of the multiple stem regions are identical to one another. In some embodiments, the nucleotide sequences of at least one of the multiple stem regions is not identical to those of the others. In some embodiments, the guide nucleic acid comprises at least 2, at least 3, at least 4, or at least 5 stem regions.

[266] In some embodiments, the compositions, systems, and methods of the present disclosure comprise two or more guide nucleic acids (*e.g.*, 2, 3, 4, 5, 6, 7, 9, 10 or more guide nucleic acids), and/or uses thereof. In some embodiments, the compositions, systems, and methods of the present disclosure may comprise an additional guide nucleic acid or a use thereof. An additional guide nucleic acid can

target an effector protein to a different location in the target nucleic acid by binding to a different portion of the target nucleic acid from the first guide nucleic acid. For example, a guide nucleic acid can bind a portion of the target nucleic acid that is upstream or downstream of the target gene in a cell or subject as described herein, wherein the additional guide nucleic acid can bind to a portion of the target nucleic acid that is located either upstream or downstream of where the first guide RNA has targeted. In some embodiments, the first loci and the second loci of the target nucleic acid may be located at least 1, at least 10, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90 or at least 100 nucleotides apart. In some embodiments, the first loci and the second loci of the target nucleic acid may be located between 100 and 200, 200 and 300, 300 and 400, 400 and 500, 500 and 600, 600 and 700, 700 and 800, 800 and 900 or 900 and 1000 nucleotides apart. In some embodiments, the first loci and/or the second loci of the target nucleic acid are located in an intron of a gene. In some embodiments, the first loci and/or the second loci of the target nucleic acid are located in an exon of a gene. In some embodiments, the first loci and/or the second loci of the target nucleic acid span an exon-intron junction of a gene. In some embodiments, the first portion and/or the second portion of the target nucleic acid are located on either side of an exon and cutting at both sites results in deletion of the exon. In some embodiments, composition, systems, and methods comprise a donor nucleic acid that may be inserted in replacement of a deleted or cleaved sequence of the target nucleic acid. In some embodiments, compositions, systems, and methods comprising multiple guide nucleic acids or uses thereof comprise multiple effector proteins, wherein the effector proteins may be identical, non-identical, or combinations thereof.

[267] In some embodiments, the dual-guided compositions, systems, and methods described herein can modify the target nucleic acid in two locations. In some embodiments, the dual-guided compositions, systems, and methods described herein can cleave the target nucleic acid in the two locations targeted by the guide RNAs. In some embodiments, a donor nucleic acid is inserted in replacement of the deleted sequence. The modification of the target nucleic acid at two different loci is referred to herein as "dual-cutting". Accordingly, in some embodiments, dual-guide nucleic acid compositions, systems, and methods can comprise two effector proteins, individually corresponding a guide RNA or a single effector protein with two different guide RNA to achieve dual-cutting.

[268] In some embodiments, a guide nucleic acid comprises at least 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 contiguous nucleotides that are complementary to a eukaryotic sequence. Such a eukaryotic sequence is a nucleotide sequence that is present in a host eukaryotic cell. Such a nucleotide sequence is distinguished from nucleotide sequences present in other host cells, such as prokaryotic cells, or viruses. Said sequences present in a eukaryotic cell can be located in a gene, an exon, an intron, a non-coding (*e.g.*, promoter or enhancer) region, a selectable marker, tag, signal, and the like. In some embodiments, a target sequence is a eukaryotic sequence.

[269] In some embodiments, a length of a guide nucleic acid is about 30 to about 120 linked nucleotides. In some embodiments, the length of a guide nucleic acid is about 40 to about 100, about

40 to about 90, about 40 to about 80, about 40 to about 70, about 40 to about 60, about 40 to about 50, about 50 to about 90, about 50 to about 80, about 50 to about 70, or about 50 to about 60 linked nucleotides. In some embodiments, the length of a guide nucleic acid is about 40, about 45, about 50, about 55, about 60, about 65, about 70 or about 75 linked nucleotides. In some embodiments, the length of a guide nucleic acid is greater than about 15, about 20, about 25, about 30, about 35, about 40, about 45, about 50, about 55, about 60, about 65, about 70 or about 75 linked nucleotides. In some embodiments, the length of a guide nucleic acid is not greater than about 40, about 45, about 50, about 55, about 60, about 65, about 70, about 75, about 80, about 85, about 90, about 95, about 100, about 105, about 110, about 115, about 120, or about 125 linked nucleotides.

[270] In some embodiments, guide nucleic acids comprise additional elements that contribute additional functionality (*e.g.*, stability, heat resistance, *etc.*) to the guide nucleic acid. Such elements may be one or more nucleotide alterations, nucleotide sequences, intermolecular secondary structures, or intramolecular secondary structures (*e.g.*, one or more hair pin regions, one or more bulges, *etc.*).

[271] In some embodiments, guide nucleic acids comprise one or more linkers connecting different nucleotide sequences as described herein. A linker may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more nucleotides. A linker may be any suitable linker, examples of which are described herein.

[272] In some embodiments, the guide nucleic acid comprises 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 linked nucleosides. In general, a guide nucleic acid comprises at least 10 linked nucleosides. In some embodiments, a guide nucleic acid comprises at least 25 linked nucleosides. A guide nucleic acid may comprise 10 to 50 linked nucleosides. In some embodiments, the guide nucleic acid comprises or consists essentially of about 12 to about 80 linked nucleosides, about 12 to about 50, about 12 to about 45, about 12 to about 40, about 12 to about 35, about 12 to about 30, about 12 to about 25, from about 12 to about 20, about 12 to about 19 , about 19 to about 20, about 19 to about 25, about 19 to about 30, about 19 to about 35, about 19 to about 40, about 19 to about 45, about 19 to about 50, about 19 to about 60, about 20 to about 25, about 20 to about 30, about 20 to about 35, about 20 to about 40, about 20 to about 45, about 20 to about 50, or about 20 to about 60 linked nucleosides. In some embodiments, the guide nucleic acid has about 10 to about 60, about 20 to about 50, or about 30 to about 40 linked nucleosides.

[273] In some embodiments, the guide nucleic acid or a nucleic acid encoding the guide nucleic acid comprises a nucleotide sequence as described herein (*e.g.*, **TABLE 4, TABLE 5, TABLE 6, TABLE 7, TABLE 8, or TABLE 9**). Such nucleotide sequences described herein (*e.g.*, **TABLE 4, TABLE 5, TABLE 6, TABLE 7, TABLE 8, or TABLE 9**) may be described as a nucleotide sequence of either DNA or RNA, however, no matter the form the sequence is described, it is readily understood that such nucleotide sequences can be revised to be RNA or DNA, as needed, for describing a sequence within a guide nucleic acid itself or the sequence that encodes a guide nucleic acid, such as a nucleotide sequence described herein for a vector. Similarly, disclosure of the nucleotide sequences described herein (*e.g.*, **TABLE 4, TABLE 5, TABLE 6, TABLE 7, TABLE 8, or TABLE 9**) also discloses the

complementary nucleotide sequence, the reverse nucleotide sequence, and the reverse complement nucleotide sequence, any one of which can be a nucleotide sequence for use in a guide nucleic acid as described herein.

[274] In some embodiments, the guide nucleic acid or a nucleic acid encoding the guide nucleic acid comprises a sequence that is at least 70%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of the nucleotide sequences recited in **TABLE 4, TABLE 5, TABLE 6, TABLE 7, TABLE 8, TABLE 9** or any combination thereof. In some embodiments, the guide nucleic acid or a nucleic acid encoding the guide nucleic acid comprises a nucleotide sequence that is at least 70%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of **SEQ ID NO: 1825, 1828, 1846** and **1847**. In some embodiments, the guide nucleic acid comprises a nucleotide sequence that is at least 70%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of **SEQ ID NO: 652, 685, 705, 713, 777** and **790**. In some embodiments, the guide nucleic acid comprises a nucleotide sequence that is at least 70%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of **SEQ ID NO: 856, 864** and **866**.

[275] In some embodiments, the guide nucleic acid or a nucleic acid encoding the guide nucleic acid comprises a spacer sequence and/or a repeat sequence. In some embodiments, the guide nucleic acid or a nucleic acid encoding the guide nucleic acid comprises a spacer sequence and/or a handle sequence. In some embodiments, the handle sequence comprises one or more of an intermediary sequence, a tracrRNA sequence or a portion thereof, a repeat sequence, and a linker.

[276] **TABLE 13**, **TABLE 14**, **TABLE 15**, **TABLE 16**, **TABLE 17**, **TABLE 19**, **TABLE 20**, and **TABLE 21** provide exemplary compositions comprising an effector protein described herein and gRNAs (or components thereof). Each row represents an exemplary composition comprising an effector protein, PAM sequences recognized by the effector protein, and a guide nucleic acid or components thereof, wherein the guide nucleic acid interacts with the effector protein. In some embodiments, the nucleotide sequence of the guide nucleic acid is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the nucleotide sequences recited in **TABLE 13**, **TABLE 14**, **TABLE 15**, **TABLE 16**, **TABLE 17**, **TABLE 19**, **TABLE 20**, and **TABLE 21**. In some embodiments, an exemplary composition may comprise an effector protein comprising an amino acid sequence of **SEQ ID NO: 6 or 228** recognizing any one of PAM sequences recited in **TABLE 13**, and a guide nucleic acid that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the nucleotide sequences recited in **TABLE 13**. In some embodiments, an exemplary composition may comprise an effector protein comprising an amino acid sequence of **SEQ ID NO: 47** recognizing any one of PAM sequences recited in **TABLE 14**, and a guide nucleic acid that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the nucleotide sequences recited in **TABLE 14**. In some embodiments, an exemplary

composition may comprise an effector protein comprising an amino acid sequence of **SEQ ID NO: 231** recognizing any one of PAM sequences recited in **TABLE 15**, and a guide nucleic acid that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the nucleotide sequences recited in **TABLE 15**. In some embodiments, an exemplary composition may comprise an effector protein comprising an amino acid sequence of **SEQ ID NO: 228** recognizing any one of PAM sequences recited in **TABLE 16**, and a guide nucleic acid that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the nucleotide sequences recited in **TABLE 16**. In some embodiments, an exemplary composition may comprise an effector protein comprising any one of amino acid sequences of **SEQ ID NO: 6** and **228-230** recognizing any one of PAM sequences recited in **TABLE 17**, and a guide nucleic acid comprising a repeat sequence and a spacer sequence, wherein the repeat sequence is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 252** and the spacer sequence is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the spacer sequences recited in **TABLE 17**. In some embodiments, an exemplary composition may comprise an effector protein comprising an amino acid sequence of **SEQ ID NO: 13** recognizing any one of PAM sequences recited in **TABLE 19**, and a guide nucleic acid comprising a handle sequence and a spacer sequence, wherein the handle sequence is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the handle sequences recited in **TABLE 19** and the spacer sequence is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the spacer sequences recited in **TABLE 19**. In some embodiments, an exemplary composition may comprise an effector protein comprising an amino acid sequence of **SEQ ID NO: 47** recognizing any one of PAM sequences recited in **TABLE 20**, and a guide nucleic acid comprising a repeat sequence and a spacer sequence, wherein the repeat sequence is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 254** and the spacer sequence is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the spacer sequences recited in **TABLE 20**. In some embodiments, an exemplary composition may comprise an effector protein comprising any one of amino acid sequences of **SEQ ID NO: 181-184** recognizing any one of PAM sequences recited in **TABLE 21**, and a guide nucleic acid comprising a handle sequence and a spacer sequence, wherein the handle sequence is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the handle sequences recited in **TABLE 21** and the spacer sequence is at least 65%, at

least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the spacer sequences recited in **TABLE 21**.

Spacer Region

[277] Guide nucleic acids described herein may comprise one or more spacer sequences. In some embodiments, a spacer sequence is capable of hybridizing to a target sequence of a target nucleic acid. In some embodiments, a spacer sequence comprises a nucleotide sequence that is, at least partially, hybridizable to an equal length of a nucleotide sequence (*e.g.*, a target sequence) of a target nucleic acid. Exemplary hybridization conditions are described herein. In some embodiments, the spacer sequence may function to direct an RNP complex comprising the guide nucleic acid to the target nucleic acid for detection and/or modification. The spacer sequence may function to direct a RNP to the target nucleic acid for detection and/or modification. A spacer sequence may be complementary to a target sequence that is adjacent to a PAM that is recognizable by an effector protein described herein.

[278] In general, a guide nucleic acid comprises a spacer region that hybridizes to a target sequence of a target nucleic acid. The spacer region may comprise complementarity with (*e.g.*, hybridize to) a target sequence of a target nucleic acid. In some embodiments, a spacer sequence comprises at least 5 to about 50 contiguous nucleotides that are complementary to a target sequence in a target nucleic acid. In some embodiments, a spacer sequence comprises at least 5 to about 50 linked nucleotides. In some embodiments, a spacer sequence comprises at least 5 to about 50, at least 5 to about 25, at least about 10 to at least about 25, or at least about 15 to about 25 linked nucleotides. In some embodiments, the spacer region is 15-28 linked nucleosides in length. In some embodiments, the spacer region is 15-26, 15-24, 15-22, 15-20, 15-18, 16-28, 16-26, 16-24, 16-22, 16-20, 16-18, 17-26, 17-24, 17-22, 17-20, 17-18, 18-26, 18-24, or 18-22 linked nucleosides in length. In some embodiments, the spacer region is 18-24 linked nucleosides in length. In some embodiments, the spacer region is at least 15 linked nucleosides in length. In some embodiments, the spacer region is at least 16, 18, 20, or 22 linked nucleosides in length. In some embodiments, the spacer region comprises at least 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides. In some embodiments, the spacer region is at least 17 linked nucleosides in length. In some embodiments, the spacer region is at least 18 linked nucleosides in length. In some embodiments, the spacer region is at least 20 linked nucleosides in length. In some embodiments, the spacer region is at least 80 %, at least 85 %, at least 90 %, at least 95 % or 100 % complementary to a target sequence of the target nucleic acid. % Complementary is a percent of nucleotides in two equal length nucleotide sequences that can undergo cumulative base pairing at two or more individual corresponding positions in an antiparallel orientation. Accordingly, the nucleic acid sequences that are not completely complementary over their entire length comprises one or more mismatches. A mismatch may present at any position in the two opposed nucleotides that are not complementary. The % complementary is calculated by dividing the total number of the complementary residues by the total number of the nucleotides in one of the equal length sequences, and multiplying

by 100. Complete or total complementarity nucleotide sequence comprises 100% of the residues of the nucleotide sequence are complementary to residues in a reference nucleotide sequence. Partially complementarity nucleotide sequence comprises at least 20%, but less than 100%, of the residues of the nucleotide sequence are complementary to residues in a reference nucleotide sequence. In some embodiments, at least 50%, but less than 100%, of the residues of a nucleotide sequence are complementary to residues in a reference nucleotide sequence. In some embodiments, at least 70%, 80%, 90% or 95%, but less than 100%, of the residues of a nucleotide sequence are complementary to residues in a reference nucleotide sequence. Noncomplementary nucleotide sequences comprises less than 20% of the residues of the nucleotide sequence are complementary to residues in a reference nucleotide sequence. In some embodiments, the spacer region is 100 % complementary to the target sequence of the target nucleic acid. In some embodiments, the spacer region comprises at least 15 contiguous nucleotides that are complementary to the target nucleic acid. In some embodiments, the spacer region comprises at least 17 contiguous nucleotides that are complementary to the target nucleic acid.

[279] In some embodiments, a spacer sequence is adjacent to a repeat sequence. In some embodiments, a spacer sequence follows a repeat sequence in a 5' to 3' direction. In some embodiments, a spacer sequence precedes a repeat sequence in a 5' to 3' direction. In some embodiments, the spacer sequence(s) and the repeat sequence(s) of the guide nucleic acid are present within the same molecule. In some embodiments, the spacer(s) and repeat sequence(s) are linked directly to one another. In some embodiments, a linker is present between the spacer(s) and repeat sequences. Linkers may be any suitable linker. In some embodiments, the spacer sequence(s) and the repeat sequence(s) of the guide nucleic acid are present in separate molecules, which are joined to one another by base pairing interactions.

[280] It is understood that the nucleotide sequence of a spacer region need not be 100 % complementary to that of a target sequence of a target nucleic acid to hybridize or hybridize specifically to the target sequence. The guide nucleic acid may comprise at least one uracil between nucleic acid residues 5 to 20 of the spacer region that is not complementary to the corresponding nucleoside of the target sequence. The guide nucleic acid may comprise at least one uracil between nucleic acid residues 5 to 9, 10 to 14, or 15 to 20 of the spacer region that is not complementary to the corresponding nucleoside of the target sequence. In some embodiments, a spacer sequence comprises a nucleotide sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% or 100% complementary to a target sequence of a target nucleic acid. A spacer sequence is capable of hybridizing to an equal length portion of a target nucleic acid (*e.g.*, a target sequence). In some embodiments, the region of the target nucleic acid that is complementary to the spacer region comprises an epigenetic modification or a post-transcriptional modification. In some embodiments, the epigenetic modification comprises an acetylation, methylation, or thiol modification.

[281] In some embodiments, a target nucleic acid comprises any one of the nucleotide sequences recited in **TABLE 10**. In some embodiments, a spacer sequence is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% or 100% complementary to a target sequence of any one of target nucleic acids recited in **TABLE 10**. For example, the spacer sequence may comprise at least one modification, such as substituted or modified nucleotide, that is not complementary to the corresponding nucleotide of the target sequence. In some embodiments, the spacer sequence comprises at least 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 contiguous nucleotides that are capable of hybridizing to the target sequence. In some embodiments, the spacer sequence comprises at least 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 contiguous nucleotides that are complementary to the target sequence. In some embodiments, a spacer region is capable of hybridizing to a target sequence within a target nucleic acid, wherein the target nucleic acid is a safe harbor. In some embodiments, the spacer region is capable of hybridizing to a target sequence within a safe harbor loci. In some embodiments, the safe harbor locus can be any one of the loci of *AAVS1* (PPP1R12C), *ALB*, *Angptl3*, *ApoC3*, *ASGR2*, *CCR5*, *FIX* (F9), *G6PC*, *Gys2*, *HGD*, *Lp*(a), *Pcsk9*, *Serpina1*, *TF*, or *TTR*. In some embodiments, the guide nucleic acid comprises a spacer sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% or 100% identical to any one of the nucleotide sequences recited in **TABLE 4** and **TABLE 5**.

[282] **TABLE 4** provides illustrative spacer sequences for use with the compositions, systems and methods of the disclosure. In some embodiments, the spacer sequence comprises a nucleotide sequence that is at least 70%, at least 80%, at least 90%, at least 92%, at least 95%, at least 97%, at least 99%, or 100% identical to any one of the nucleotide sequences recited in **TABLE 4**, wherein the spacer sequence hybridizes to a target sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to **SEQ ID NO: 266**. In some embodiments, the spacer sequence comprises an RNA sequence that is at least 70%, at least 80%, at least 90%, at least 92%, at least 95%, at least 97%, at least 99%, or 100% identical to any one of **SEQ ID NO: 301-604**. In some embodiments, the spacer sequence comprises an RNA sequence that is at least 70%, at least 80%, at least 90%, at least 92%, at least 95%, at least 97%, at least 99%, or 100% identical to any one of **SEQ ID NO: 348, 558, 401, 409, 473** and **486**. In some embodiments, the spacer sequence comprises an RNA sequence that is at least 70%, at least 80%, at least 90%, at least 92%, at least 95%, at least 97%, at least 99%, or 100% identical to any one of **SEQ ID NO: 592, 600** and **602**. In some embodiments, wherein the spacer sequence comprises a nucleotide sequence that is at least 70%, at least 80%, at least 90%, at least 92%, at least 95%, at least 97%, at least 99%, or 100% identical to any one of nucleotide sequences recited in **SEQ ID NO: 551-575**.

[283] In some embodiments, the spacer sequence comprises one or more nucleobase alterations at one or more positions in any one of the nucleotide sequences recited in **TABLE 4**. Alternative nucleobases can be any one or more of A, C, G, T or U, or a deletion, or an insertion.

[284] In some embodiments, a target nucleic acid is a safe harbor within human *SERPINA1* gene (**SEQ ID NO: 1795**). In some embodiments, the safe harbor comprises exon 1 of human *SERPINA1* gene, intron 1 of human *SERPINA1* gene, exon 2 of human *SERPINA1* gene, or a combination thereof. In some embodiments, a spacer sequence for use with the compositions, systems and methods of the disclosure comprises a nucleotide sequence that is capable of hybridizing with the safe harbor within human *SERPINA1* gene (**SEQ ID NO: 1795**). **TABLE 5** provides exemplary spacer sequences targeting intron 1 of human *SERPINA1* gene. In some embodiments, the spacer sequence comprises a nucleotide sequence that is at least 70%, at least 80%, at least 90%, at least 92%, at least 95%, at least 97%, at least 99%, or 100% identical to any one of the nucleotide sequences recited in **TABLE 5**, wherein the spacer sequence is capable of hybridizing to a target sequence within intron 1 of human *SERPINA1* gene.

[285] In some embodiments, the spacer sequence comprises one or more nucleobase alterations at one or more positions in any one of the nucleotide sequences recited in **TABLE 5**. Alternative nucleobases can be any one or more of A, C, G, T or U, or a deletion, or an insertion.

Repeat Region

[286] Guide nucleic acids described herein may comprise one or more repeat sequences. In some embodiments, a repeat sequence comprises a nucleotide sequence that is not complementary to a target sequence of a target nucleic acid. In some embodiments, a repeat sequence comprises a nucleotide sequence that may interact with an effector protein. In some embodiments, a repeat sequence is connected to another sequence of a guide nucleic acid, such as an intermediary sequence, that is capable of non-covalently interacting with an effector protein. In some embodiments, a repeat sequence includes a nucleotide sequence that is capable of forming a guide nucleic acid-effector protein complex (*e.g.*, a RNP complex).

[287] In some embodiments, the repeat sequence is between 10 and 50, 12 and 48, 14 and 46, 16 and 44, and 18 and 42 nucleotides in length.

[288] In some embodiments, a repeat sequence is adjacent to a spacer sequence. In some embodiments, a repeat sequence is followed by a spacer sequence in the 5' to 3' direction. In some embodiments, a repeat sequence is preceded by a spacer sequence in the 5' to 3' direction. In some embodiments, a repeat sequence is adjacent to an intermediary sequence. In some embodiments, a repeat sequence is 3' to an intermediary sequence. In some embodiments, an intermediary sequence is followed by a repeat sequence, which is followed by a spacer sequence in the 5' to 3' direction. In some embodiments, a repeat sequence is linked to a spacer sequence and/or an intermediary sequence. In some embodiments, a guide nucleic acid comprises a repeat sequence linked to a spacer sequence and/or to an intermediary sequence, which may be a direct link or by any suitable linker, examples of which are described herein.

[289] In some embodiments, guide nucleic acids comprise more than one repeat sequence (*e.g.*, two or more, three or more, or four or more repeat sequences). In some embodiments, a guide nucleic acid

comprises more than one repeat sequence separated by another sequence of the guide nucleic acid. For example, in some embodiments, a guide nucleic acid comprises two repeat sequences, wherein the first repeat sequence is followed by a spacer sequence, and the spacer sequence is followed by a second repeat sequence in the 5' to 3' direction. In some embodiments, the more than one repeat sequences are identical. In some embodiments, the more than one repeat sequences are not identical.

[290] In some embodiments, the repeat sequence comprises two sequences that are complementary to each other and hybridize to form a double stranded RNA duplex (dsRNA duplex). In some embodiments, the two sequences are not directly linked and hybridize to form a stem loop structure. In some embodiments, the dsRNA duplex comprises 5, 10, 15, 20 or 25 base pairs (bp). In some embodiments, not all nucleotides of the dsRNA duplex are paired, and, therefore, the duplex forming sequence may include a bulge. In some embodiments, the repeat sequence comprises a hairpin or stem-loop structure, optionally at the 5' portion of the repeat sequence. In some embodiments, a strand of the stem portion comprises a sequence and the other strand of the stem portion comprises a sequence that is, at least partially, complementary. In some embodiments, such sequences may have 65% to 100% complementarity (*e.g.*, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% complementarity). In some embodiments, a guide nucleic acid comprises nucleotide sequence that when involved in hybridization events may hybridize over one or more segments such that intervening or adjacent segments are not involved in the hybridization event (*e.g.*, a bulge, a loop structure or hairpin structure, *etc.*).

[291] In general, a guide nucleic acid comprises a repeat region that interacts with the effector protein. In some embodiments, the repeat region may also be referred to as a "protein-binding segment." Typically, the repeat region is adjacent to the spacer region. For example, a guide RNA that interacts with an effector protein comprises a repeat region that is 5' of the spacer region.

[292] **TABLE 6** provides illustrative repeat sequences for use with the compositions and methods of the disclosure. In some embodiments, the repeat sequence comprises at least 70%, at least 80%, at least 90%, at least 92%, at least 95%, at least 97%, at least 99%, or 100% sequence identity to any one of the nucleotide sequences as set forth in **TABLE 6**.

[293] In some embodiments, the repeat sequence comprises one or more nucleobase alterations at one or more positions in any one of the nucleotide sequences of **TABLE 6**. Alternative nucleobases can be any one or more of A, C, G, T or U, or a deletion, or an insertion.

[294] In some embodiments, the repeat sequence comprises an RNA sequence that is 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% identical to any one of **SEQ ID NO: 252-258, 1789** and **1848**.

[295] In some embodiments, compositions, systems and methods described herein comprise a crRNA sequence that is at least 65%, at least 70%, at least 80%, at least 90%, at least 92%, at least 95%, at least 97%, at least 99%, or 100% identical to any one of **SEQ ID NO: 252-258, 1789** and **1848**. In some embodiments, compositions, systems and methods described herein comprise a crRNA comprising a repeat sequence that is 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% identical to any one of **SEQ ID NO: 252-258, 1789** and **1848**.

[296] In some embodiments, compositions disclosed herein comprises an effector protein comprising an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% identical to any one of **SEQ ID NO: 6-12,** and **228- 230**; a crRNA comprising a repeat sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to **SEQ ID NO: 252** or **1848**.

[297] In some embodiments, compositions disclosed herein comprises an effector protein comprising an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% identical to **SEQ ID NO: 13**; a crRNA comprising a repeat sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to **SEQ ID NO: 253**.

[298] In some embodiments, compositions disclosed herein comprises an effector protein comprising an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% identical to **SEQ ID NO: 47**; a crRNA comprising a repeat sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to **SEQ ID NO: 254**.

[299] In some embodiments, compositions disclosed herein comprises an effector protein comprising an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% identical to **SEQ ID NO: 231**; a crRNA comprising a repeat sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to **SEQ ID NO: 255** or **1789**.

[300] In some embodiments, compositions disclosed herein comprises an effector protein comprising an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% identical to any one of **SEQ ID NO: 181-184**; a crRNA comprising a repeat sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to **SEQ ID NO: 256**.

[301] In some embodiments, compositions disclosed herein comprises an effector protein comprising an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% identical to **SEQ ID NO: 187** or **188**; a crRNA comprising a repeat sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to **SEQ ID NO: 257**.

[302] In some embodiments, compositions disclosed herein comprises an effector protein comprising an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% identical to any one of **SEQ ID NO: 181-185**; a crRNA comprising a repeat sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to **SEQ ID NO: 258**.

Intermediary sequence

[303] Guide nucleic acids described herein may comprise one or more intermediary sequences. In general, an intermediary sequence used in the present disclosure is not transactivated or transactivating. An intermediary sequence may also be referred to as an intermediary RNA, although it may comprise deoxyribonucleotides instead of or in addition to ribonucleotides, and/or modified bases. In general, the intermediary sequence non-covalently binds to an effector protein. In some embodiments, the intermediary sequence forms a secondary structure, for example in a cell, and an effector protein binds the secondary structure.

[304] In some embodiments, a length of the intermediary sequence is at least 30, 50, 70, 90, 110, 130, 150, 170, 190, or 210 linked nucleotides. In some embodiments, a length of the intermediary sequence is not greater than 30, 50, 70, 90, 110, 130, 150, 170, 190, or 210 linked nucleotides. In some embodiments, the length of the intermediary sequence is about 30 to about 210, about 60 to about 210, about 90 to about 210, about 120 to about 210, about 150 to about 210, about 180 to about 210, about 30 to about 180, about 60 to about 180, about 90 to about 180, about 120 to about 180, or about 150 to about 180 linked nucleotides.

[305] An intermediary sequence may also comprise or form a secondary structure (*e.g.*, one or more hairpin loops) that facilitates the binding of an effector protein to a guide nucleic acid and/or modification activity of an effector protein on a target nucleic acid (*e.g.*, a hairpin region). An intermediary sequence may comprise from 5' to 3', a 5' region, a hairpin region, and a 3' region. In some embodiments, the 5' region may hybridize to the 3' region. In some embodiments, the 5' region of the intermediary sequence does not hybridize to the 3' region.

[306] In some embodiments, the hairpin region may comprise a first sequence, a second sequence that is reverse complementary to the first sequence, and a stem-loop linking the first sequence and the second sequence. In some embodiments, an intermediary sequence comprises a stem-loop structure comprising a stem region and a loop region. In some embodiments, the stem region is 4 to 8 linked nucleotides in length. In some embodiments, the stem region is 5 to 6 linked nucleotides in length. In some embodiments, the stem region is 4 to 5 linked nucleotides in length. In some embodiments, an intermediary sequence comprises a pseudoknot (*e.g.*, a secondary structure comprising a stem at least partially hybridized to a second stem or half-stem secondary structure). An effector protein may interact with an intermediary sequence comprising a single stem region or multiple stem regions. In some embodiments, the nucleotide sequences of the multiple stem regions are identical to one another. In some embodiments, the nucleotide sequences of at least one of the multiple stem regions is not identical to those of the others. In some embodiments, an intermediary sequence comprises 1, 2, 3, 4, 5 or more stem regions.

[307] In some embodiments, an intermediary sequence comprises a nucleotide sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, or at least 98%, at least 99%, or 100% identical to any one of the intermediary sequences in

TABLE 7. In some embodiments, an intermediary sequence comprises at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, 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 110, at least 120, at least 130, or at least 140 contiguous nucleotides of any one of the intermediary sequences recited in **TABLE 7.**

Handle sequence

[308] In some embodiments, compositions, systems and methods described herein comprise the nucleic acid, wherein the nucleic acid comprises a handle sequence. In some embodiments, the handle sequence comprises an intermediary sequence. In some embodiments, the intermediary sequence is at the 3'-end of the handle sequence. In some embodiments, the intermediary sequence is at the 5'- end of the handle sequence. In some embodiments, the handle sequence further comprises one or more of linkers and repeat sequences. In some embodiments, the linker comprises a sequence of 5'-GAAA-3.' In some embodiments, the intermediary sequence is 5' to the repeat sequence. In some embodiments, the intermediary sequence is 5' to the linker. In some embodiments, the intermediary sequence is 3' to the repeat sequence. In some embodiments, the intermediary sequence is 3' to the linker. In some embodiments, the repeat sequence is 3' to the linker. In some embodiments, the repeat sequence is 5' to the linker.

[309] In some embodiments, the handle sequence is: 1) capable of being non-covalently bound by an effector protein and 2) connects the portion of the sgRNA capable of being non-covalently bound by an effector protein to a nucleotide sequence that is hybridizable to a target nucleic acid. In general, the handle sequence comprises an intermediary sequence, that is capable of being non-covalently bound by an effector protein. In some embodiments, the handle sequence further comprises a repeat sequence. In such embodiments, the intermediary sequence or a combination of the intermediary sequence and the repeat sequence is capable of being non-covalently bound by an effector protein.

[310] In some embodiments, a sgRNA may include a handle sequence having a hairpin region, as well as a linker and a repeat sequence. The sgRNA having a handle sequence can have a hairpin region positioned 3' of the linker and/or repeat sequence. The sgRNA having a handle sequence can have a hairpin region positioned 5' of the linker and/or repeat sequence. The hairpin region may include a first nucleotide sequence, a second nucleotide sequence that is reverse complementary to the first nucleotide sequence, and a stem-loop linking the first nucleotide sequence and the second nucleotide sequence.

[311] In some embodiments, an effector protein may recognize a secondary structure of a handle sequence. In some embodiments, at least a portion of the handle sequence interacts with an effector protein described herein. Accordingly, in some embodiments, at least a portion of the intermediary sequence interacts with the effector protein described herein. In some embodiments, both, at least a portion of the intermediary sequence and at least a portion of the repeat sequence, interacts with the effector protein. In general, the handle sequence is capable of interacting (*e.g.*, non-covalent binding) with any one of the effector proteins described herein.

[312] In some embodiments, the handle sequence of a sgRNA comprises a stem-loop structure comprising a stem region and a loop region. In some embodiments, the stem region is 4 to 8 linked nucleotides in length. In some embodiments, the stem region is 5 to 6 linked nucleotides in length. In some embodiments, the stem region is 4 to 5 linked nucleotides in length. In some embodiments, the sgRNA comprises a pseudoknot (*e.g.*, a secondary structure comprising a stem at least partially hybridized to a second stem or half-stem secondary structure). An effector protein may recognize a sgRNA comprising multiple stem regions. In some embodiments, the nucleotide sequences of the multiple stem regions are identical to one another. In some embodiments, the nucleotide sequences of at least one of the multiple stem regions is not identical to those of the others. In some embodiments, the sgRNA comprises at least 2, at least 3, at least 4, or at least 5 stem regions.

[313] A handle sequence may include deoxyribonucleosides, ribonucleosides, chemically modified nucleosides, or any combination thereof. In some embodiments, a length of the handle sequence is at least 30, 50, 70, 90, 110, 130, 150, 170, 190, or 210 linked nucleotides. In some embodiments, a length of the handle sequence is not greater than 30, 50, 70, 90, 110, 130, 150, 170, 190, or 210 linked nucleotides. In some embodiments, the length of the handle sequence is about 30 to about 210, about 60 to about 210, about 90 to about 210, about 120 to about 210, about 150 to about 210, about 180 to about 210, about 30 to about 180, about 60 to about 180, about 90 to about 180, about 120 to about 180, or about 150 to about 180 linked nucleotides.

[314] In some embodiments, the length of a handle sequence in a sgRNA is not greater than 50, 56, 66, 67, 68, 69, 70, 71, 72, 73, 95, or 105 linked nucleotides. In some embodiments, the length of a handle sequence in a sgRNA is about 30 to about 120 linked nucleotides. In some embodiments, the length of a handle sequence in a sgRNA is about 50 to about 105, about 50 to about 95, about 50 to about 73, about 50 to about 71, about 50 to about 70, or about 50 to about 69 linked nucleotides. In some embodiments, the length of a handle sequence in a sgRNA is 56 to 105 linked nucleotides, from 56 to 105 linked nucleotides, 66 to 105 linked nucleotides, 67 to 105 linked nucleotides, 68 to 105 linked nucleotides, 69 to 105 linked nucleotides, 70 to 105 linked nucleotides, 71 to 105 linked nucleotides, 72 to 105 linked nucleotides, 73 to 105 linked nucleotides, or 95 to 105 linked nucleotides. In some embodiments, the length of a handle sequence in a sgRNA is 40 to 70 nucleotides. In some embodiments, the length of a handle sequence in a sgRNA is 50, 56, 66, 67, 68, 69, 70, 71, 72, 73, 95, or 105 linked nucleotides. In some embodiments, the length of a handle sequence in a sgRNA is 69 nucleotides.

[315] **TABLE 7** provides illustrative handle sequence for an sgRNA and exemplary portions of a sgRNA (a handle sequence without a linker or repeat sequence) for use with the compositions and methods of the disclosure. In some embodiments, the handle sequence comprises a sequence that is at least 65%, at least 70%, at least 80%, at least 90%, at least 92%, at least 95%, at least 97%, at least 99%, or 100% identical to any one of nucleotide sequences recited in **TABLE 7**. In some embodiments, the handle sequence comprises at least 5, at least 10, 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 110, at least 120, at least 130, or at least 140 contiguous nucleotides of any one of the nucleotide sequences recited in **TABLE 7.**

[316] In some embodiments, the handle sequence comprises an RNA sequence that is at least 70%, at least 80%, at least 90%, at least 92%, at least 95%, at least 97%, at least 99%, or 100% identical to any one of **SEQ ID NO: 259-264**.

Linker For Nucleic Acids

[317] In some embodiments, a guide nucleic acid for use with compositions, systems, and methods described herein comprises one or more linkers, or a nucleic acid encoding one or more linkers. In some embodiments, the guide nucleic acid comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, or at least ten linkers. In some embodiments, the guide nucleic acid comprises one, two, three, four, five, six, seven, eight, nine, or ten linkers. In some embodiments, the guide nucleic acid comprises more than one linker. In some embodiments, at least two of the more than one linker are same. In some embodiments, at least two of the more than one linker are not same.

[318] In some embodiments, a linker comprises one to ten, one to seven, one to five, one to three, two to ten, two to eight, two to six, two to four, three to ten, three to seven, three to five, four to ten, four to eight, four to six, five to ten, five to seven, six to ten, six to eight, seven to ten, or eight to ten linked nucleotides. In some embodiments, the linker comprises one, two, three, four, five, six, seven, eight, nine, or ten linked nucleotides. In some embodiments, a linker comprises a nucleotide sequence of 5'-GAAA-3'.

[319] In some embodiments, a guide nucleic acid comprises one or more linkers connecting one or more of a repeat sequence, a spacer sequence, a handle sequence, and an intermediary sequence. In some embodiments, the guide nucleic acid comprises one or more linkers connecting one or more of: a repeat sequence and a spacer sequence; a handle sequence and a spacer sequence; an intermediary sequence and a repeat sequence; and an intermediary sequence and a spacer sequence. In some embodiments, the guide nucleic acid comprises at least two repeat sequences connected by a linker.

[320] In some embodiments, compositions, systems and methods described herein comprise the nucleic acid, wherein the nucleic acid comprises a handle sequence, wherein the handle sequence comprises a linker sequence. In some embodiments, the linker sequence comprises any one of the sequences recited in **TABLE 8**.

III. A Single Nucleic Acid System

[321] In some embodiments, compositions, systems and methods described herein comprise a single nucleic acid system comprising a guide nucleic acid or a nucleotide sequence encoding the guide nucleic acid, and one or more effector proteins or a nucleotide sequence encoding the one or more effector proteins. In some embodiments, a first region (FR) of the guide nucleic acid non-covalently interacts

with the one or more polypeptides described herein. In some embodiments, a second region (SR) of the guide nucleic acid hybridizes with a target sequence of the target nucleic acid. In the single nucleic acid system having a complex of the guide nucleic acid and the effector protein, the effector protein is not transactivated by the guide nucleic acid. In other words, activity of effector protein does not require binding to a second non-target nucleic acid molecule. An exemplary guide nucleic acid for a single nucleic acid system is a crRNA or a sgRNA.

[322] In some embodiments, the guide nucleic acid comprises a nucleotide sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of the nucleotide sequences recited in **TABLE 9**, wherein an effector protein has an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of **SEQ ID NO: 6-12, 228-231**, **47**, and **13**.

[323] In some embodiments, the guide nucleic acid comprises a nucleotide sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of **SEQ ID NO: 605-814**, wherein an effector protein has an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of **SEQ ID NO: 6-12,** and **228-230**. In some embodiments, the guide nucleic acid comprises a nucleotide sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of **SEQ ID NO: 815-839**, wherein an effector protein has an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 47**. In some embodiments, the guide nucleic acid comprises a nucleotide sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of **SEQ ID NO: 840-868**, wherein an effector protein has an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 231**. In some embodiments, the guide nucleic acid comprises a nucleotide sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of **SEQ ID NO: 869-908**, wherein an effector protein has an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 13**.

crRNA

[324] In some embodiments, a guide nucleic acid may comprise a CRISPR RNA (crRNA). In some embodiments, the guide nucleic acid is the crRNA. In general, a crRNA comprises a first region (FR) and a second region (SR), wherein the FR of the crRNA comprises a repeat sequence, and the SR of the crRNA comprises a spacer sequence. In some embodiments, the repeat sequence and the spacer

sequences are directly connected to each other (*e.g.*, covalent bond (phosphodiester bond)). In some embodiments, the repeat sequence and the spacer sequence are connected by a linker.

[325] In some embodiments, a crRNA is useful as a single nucleic acid system for compositions, methods, and systems described herein or as part of a single nucleic acid system for compositions, methods, and systems described herein. In some embodiments, the crRNA of the guide nucleic acid comprises a repeat region and a spacer region, wherein the repeat region binds to the effector protein and the spacer region hybridizes to a target sequence of the target nucleic acid. The repeat sequence of the crRNA may interact with an effector protein, allowing for the guide nucleic acid and the effector protein to form an RNP complex.

[326] In some embodiments, a crRNA is useful as part of a single nucleic acid system for compositions, methods, and systems described herein. In such embodiments, a single nucleic acid system comprises a guide nucleic acid comprising a crRNA wherein, a repeat sequence of a crRNA is capable of connecting a crRNA to an effector protein. In some embodiments, a single nucleic acid system comprises a guide nucleic acid comprising a crRNA linked to another nucleotide sequence that is capable of being non-covalently bond by an effector protein. In such embodiments, a repeat sequence of a crRNA can be linked to an intermediary sequence. In some embodiments, a single nucleic acid system comprises a guide nucleic acid comprising a crRNA and an intermediary sequence.

[327] A crRNA may include deoxyribonucleosides, ribonucleosides, chemically modified nucleosides, or any combination thereof. In some embodiments, a crRNA comprises about: 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, 51, 52, 53, 54, 55, 56, 57, 58, 59, or 60 linked nucleotides. In some embodiments, a crRNA comprises at least: 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60 linked nucleotides. In some embodiments, the length of the crRNA is about 20 to about 120 linked nucleotides. In some embodiments, the length of a crRNA is about 20 to about 100, about 30 to about 100, about 40 to about 100, about 40 to about 90, about 40 to about 80, about 40 to about 70, about 40 to about 60, about 40 to about 50, about 50 to about 90, about 50 to about 80, about 50 to about 70, or about 50 to about 60 linked nucleotides. In some embodiments, the length of a crRNA is about 20, about 25, about 30, about 35, about 40, about 45, about 50, about 55, about 60, about 65, about 70 or about 75 linked nucleotides.

[328] A crRNA may be the product of processing of a longer precursor CRISPR RNA (pre-crRNA) transcribed from the CRISPR array by cleavage of the pre-crRNA within each direct repeat sequence to afford shorter, mature crRNAs. A crRNA may be generated by a variety of mechanisms, including the use of dedicated endonucleases (*e.g.*, Cas6 or Cas5d in Type I and III systems), coupling of a host endonuclease (*e.g.*, RNase III) with tracrRNA (Type II systems), or a ribonuclease activity endogenous to the effector protein itself (*e.g.*, Cpfl, from Type V systems). A crRNA may also be specifically generated outside of processing of a pre-crRNA and individually contacted to an effector protein in vivo or in vitro.

[329] Exemplary crRNA sequences are recited in **TABLE 9.** In some embodiments, a crRNA or a nucleotide encoding the crRNA comprises a nucleotide sequence that is at least 65%, at least 70%, at least 80%, at least 90%, at least 92%, at least 95%, at least 97%, at least 99%, or 100% identical to any one of the crRNA sequences recited in **TABLE 9.** In some embodiments, a crRNA or a nucleotide encoding the crRNA comprises a nucleotide sequence that is at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 contiguous nucleotides of any one of the crRNA sequences recited in **TABLE 9.**

[330] In some embodiments, the guide nucleic acid is a crRNA, wherein the crRNA has a nucleotide sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of **SEQ ID NO: 605-839**, wherein an effector protein has an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of **SEQ ID NO: 6- 12, SEQ ID NO: 228-230**, and **SEQ ID NO: 47**.

sgRNA

[331] In some embodiments, a guide nucleic acid comprises a single guide RNA (sgRNA). In some embodiments, the guide nucleic acid is a sgRNA. The combination of a spacer sequence (*e.g.*, a nucleotide sequence that hybridizes to a target sequence in a target nucleic acid) with a handle sequence may be referred to herein as a single guide RNA (sgRNA), wherein the spacer sequence and the handle sequence are covalently linked. In some embodiments, the spacer sequence and handle sequence are linked by a phosphodiester bond. In some embodiments, the spacer sequence and handle sequence are linked by one or more linked nucleotides. In some embodiments, a guide nucleic acid may comprise a spacer sequence, a repeat sequence, or handle sequence, or a combination thereof. In some embodiments, the handle sequence may comprise a portion of, or all of, a repeat sequence. In general, a sgRNA comprises a first region (FR) and a second region (SR), wherein the FR comprises a handle sequence and the SR comprises a spacer sequence.

[332] In some embodiments, the compositions comprising a guide RNA and an effector protein without a tracrRNA (*e.g.*, a single nucleic acid system), wherein the guide RNA is a sgRNA. A sgRNA may include deoxyribonucleosides, ribonucleosides, chemically modified nucleosides, or any combination thereof. A sgRNA may also include a nucleotide sequence that forms a secondary structure (*e.g.*, one or more hairpin loops) that facilitates the binding of an effector protein to the sgRNA and/or modification activity of an effector protein on a target nucleic acid (*e.g.*, a hairpin region). Such a nucleotide sequence can be contained within a handle sequence as described herein.

[333] In some embodiments, a sgRNA comprises one or more of one or more of a handle sequence, an intermediary sequence, a crRNA, a repeat sequence, a spacer sequence, a linker, or combinations thereof. For example, a sgRNA comprises a handle sequence and a spacer sequence; an intermediary sequence and an crRNA; an intermediary sequence, a repeat sequence and a spacer sequence; and the like.

[334] In some embodiments, a sgRNA comprises an intermediary sequence and an crRNA. In some embodiments, an intermediary sequence is 5' to a crRNA in an sgRNA. In some embodiments, a sgRNA comprises a linked intermediary sequence and crRNA. In some embodiments, an intermediary sequence and a crRNA are linked in an sgRNA directly (*e.g.*, covalently linked, such as through a phosphodiester bond) In some embodiments, an intermediary sequence and a crRNA are linked in an sgRNA by any suitable linker, examples of which are provided herein.

[335] In some embodiments, a sgRNA comprises a handle sequence and a spacer sequence. In some embodiments, a handle sequence is 5' to a spacer sequence in an sgRNA. In some embodiments, a sgRNA comprises a linked handle sequence and spacer sequence. In some embodiments, a handle sequence and a spacer sequence are linked in an sgRNA directly (*e.g.*, covalently linked, such as through a phosphodiester bond) In some embodiments, a handle sequence and a spacer sequence are linked in an sgRNA by any suitable linker, examples of which are provided herein.

[336] In some embodiments, a sgRNA comprises an intermediary sequence, a repeat sequence, and a spacer sequence. In some embodiments, an intermediary sequence is 5' to a repeat sequence in an sgRNA. In some embodiments, a sgRNA comprises a linked intermediary sequence and repeat sequence. In some embodiments, an intermediary sequence and a repeat sequence are linked in an sgRNA directly (*e.g.*, covalently linked, such as through a phosphodiester bond) In some embodiments, an intermediary sequence and a repeat sequence are linked in an sgRNA by any suitable linker, examples of which are provided herein. In some embodiments, a repeat sequence is 5' to a spacer sequence in an sgRNA. In some embodiments, a sgRNA comprises a linked repeat sequence and spacer sequence. In some embodiments, a repeat sequence and a spacer sequence are linked in an sgRNA directly (e.g, covalently linked, such as through a phosphodiester bond) In some embodiments, a repeat sequence and a spacer sequence are linked in an sgRNA by any suitable linker, examples of which are provided herein.

[337] An exemplary handle sequence in a sgRNA may comprise, from 5' to 3', a 5' region, a hairpin region, and a 3' region. In some embodiments, the 5' region may hybridize to the 3' region. In some embodiments, the 5' region does not hybridize to the 3' region. In some embodiments, the 3' region is covalently linked to a spacer sequence (*e.g.*, through a phosphodiester bond). In some embodiments, the 5' region is covalently linked to a spacer sequence (*e.g.*, through a phosphodiester bond).

[338] In some embodiments, the sgRNA sequence comprises a nucleotide sequence that is at least 65%, at least 70%, at least 80%, at least 90%, at least 92%, at least 95%, at least 97%, at least 99%, or 100% identical to any one of the nucleotide sequences recited in **TABLE 7**.

[339] In some embodiments, the compositions, systems and methods described herein comprise a handle sequence, wherein the handle sequence further comprises a linker. In some embodiments, the linker comprises any one of the nucleotide sequence recited in **TABLE 8**. In some embodiments, the linker comprises a sequence of **SEQ ID NO: 265**.

[340] In some embodiments, compositions, systems and methods described herein comprise sgRNA, wherein the sgRNA comprises a spacer sequence.

[341] In some embodiments, compositions, methods and systems described herein comprise a repeat sequence. In some embodiments, the repeat sequence comprises an RNA sequence that is 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% identical to any one of **SEQ ID NO: 252-258, 1789** and **1848**.

[342] In some embodiments, compositions disclosed herein comprises an effector protein comprising an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% identical to **SEQ ID NO: 13**; a guide nucleic acid comprising a handle sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of **SEQ ID NO: 259-261**.

[343] In some embodiments, compositions disclosed herein comprises an effector protein comprising an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% identical to **SEQ ID NO: 231**; a guide nucleic acid comprising a handle sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of **SEQ ID NO: 262-264**.

[344] In some embodiments, the guide nucleic acid is a single guide RNA (sgRNA). Accordingly, in some embodiments, guide nucleic acids comprise a portion or all of any one of nucleotide sequences of **SEQ ID NO: 252-264,** or **1789,** and **1848**.

[345] In some embodiments, the guide nucleic acid is a sgRNA, wherein the sgRNA has a nucleotide sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of **SEQ ID NO: 840-908**, wherein an effector protein has an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of **SEQ ID NO: 231**, and **13**.

IV. A Dual Nucleic Acid System

[346] In some embodiments, compositions, systems and methods described herein comprise a dual nucleic acid system comprising a crRNA or a nucleotide sequence encoding the crRNA, a tracrRNA or a nucleotide sequence encoding the tracrRNA, and one or more effector protein or a nucleotide sequence encoding the one or more effector protein, wherein the crRNA and the tracrRNA are separate, unlinked molecules, wherein a repeat hybridization region of the tracrRNA is capable of hybridizing with an equal length portion of the crRNA to form a tracrRNA-crRNA duplex, wherein the equal length portion of the crRNA does not include a spacer sequence of the crRNA, and wherein the spacer sequence is capable of hybridizing to a target sequence of the target nucleic acid. In the dual nucleic acid system having a complex of the guide nucleic acid, tracrRNA, and the effector protein, the effector protein is transactivated by the tracrRNA. In other words, activity of effector protein requires binding to a

tracrRNA molecule. In dual nucleic acid system, a tracrRNA and/or tracrRNA-crRNA duplex may form a secondary structure that facilitates the binding of an effector protein to a tracrRNA or a tracrRNA-crRNA. In some embodiments, the secondary structure modifies activity of the effector protein on a target nucleic acid.

tracrRNA

[347] In some embodiments, a transactivating or transactivated nucleic acid (*e.g.*, trans-activating RNA, transactivating RNA and tracrRNA) comprises a first sequence that is capable of being non-covalently bound by an effector protein and a second sequence that is capable of hybridizing, at least partially, to a portion of a crRNA to form a tracrRNA-crRNA duplex in a dual nucleic acid system. The second sequence that hybridizes to a portion of a crRNA may be referred to as a repeat hybridization sequence. Accordingly, in some embodiments, a transactivating or transactivated nucleic acid in a dual nucleic acid system is capable of hybridizing, at least partially, to a crRNA to form a tracrRNA-crRNA duplex, and of interacting with an effector protein to form a complex (*e.g.*, an RNP complex). In some embodiments, a repeat hybridization sequence is a nucleotide sequence of a tracrRNA that is capable of hybridizing to a repeat sequence of a guide nucleic acid in the dual nucleic acid system.

[348] In some embodiments, the compositions comprising a guide RNA and an effector protein (*e.g.*, in a dual nucleic acid system) comprises a tracrRNA. A tracrRNA may include deoxyribonucleosides, ribonucleosides, chemically modified nucleosides, or any combination thereof. A tracrRNA may be separate from, but form a complex with, a guide nucleic acid and an effector protein. A tracrRNA may include a nucleotide sequence that hybridizes with a portion of a guide nucleic acid (*e.g.*, a repeat hybridization region). A tracrRNA may also form a secondary structure (*e.g.*, one or more hairpin loops) that facilitates the binding of an effector protein to a guide nucleic acid and/or modification activity of an effector protein on a target nucleic acid (*e.g.*, a hairpin region). A tracrRNA may include a repeat hybridization region and a hairpin region. The repeat hybridization region may hybridize to all or part of the repeat sequence of a guide nucleic acid. The repeat hybridization region may be positioned 3' of the hairpin region. The hairpin region may include a first sequence, a second sequence that is reverse complementary to the first sequence, and a stem-loop linking the first sequence and the second sequence.

[349] In some embodiments, tracrRNAs comprise a stem-loop structure comprising a stem region and a loop region. In some embodiments, the stem region is 4 to 8 linked nucleotides in length. In some embodiments, the stem region is 5 to 6 linked nucleotides in length. In some embodiments, the stem region is 4 to 5 linked nucleotides in length. In some embodiments, the tracrRNA comprises a pseudoknot (*e.g.*, a secondary structure comprising a stem at least partially hybridized to a second stem or half-stem secondary structure). An effector protein may recognize a tracrRNA comprising multiple stem regions. In some embodiments, the nucleotide sequences of the multiple stem regions are identical to one another. In some embodiments, the nucleotide sequences of at least one of the multiple stem

regions is not identical to those of the others. In some embodiments, the tracrRNA comprises at least 2, at least 3, at least 4, or at least 5 stem regions.

[350] In some embodiments, the length of a tracrRNA is not greater than 50, 56, 68, 71, 73, 95, or 105 linked nucleotides. In some embodiments, the length of a tracrRNA is about 30 to about 120 linked nucleotides. In some embodiments, the length of a tracrRNA is about 50 to about 105, about 50 to about 95, about 50 to about 73, about 50 to about 71, about 50 to about 68, or about 50 to about 56 linked nucleotides. In some embodiments, the length of a tracrRNA is 56 to 105 linked nucleotides, from 56 to 105 linked nucleotides, 68 to 105 linked nucleotides, 71 to 105 linked nucleotides, 73 to 105 linked nucleotides, or 95 to 105 linked nucleotides. In some embodiments, the length of a tracrRNA is 40 to 60 nucleotides. In some embodiments, the length of a tracrRNA is 50, 56, 68, 71, 73, 95, or 105 linked nucleotides. In some embodiments, the length of a tracrRNA is 50 nucleotides.

[351] An exemplary tracrRNA may comprise, from 5' to 3', a 5' region, a hairpin region, a repeat hybridization region, and a 3' region. In some embodiments, the 5' region may hybridize to the 3' region. In some embodiments, the 5' region does not hybridize to the 3' region. In some embodiments, a tracrRNA may comprise an un-hybridized region at the 3' end of the tracrRNA. The un-hybridized region may have a length of about 1, about 2, about 3, about 4, about 5, about 6, about 7, about 8, about 9, about 10, about 12, about 14, about 16, about 18, or about 20 linked nucleotides. In some embodiments, the length of the un-hybridized region is 0 to 20 linked nucleotides.

[352] In some embodiments, the composition comprises a tracrRNA. In some embodiments, a crRNA and tracrRNA function as two separate, unlinked molecules. In some embodiments, the composition does not comprise a tracrRNA. In some embodiments, the composition comprising an effector protein and a guide RNA does not comprise a tracrRNA. In some embodiments, an effector protein does not require a tracrRNA to locate and/or cleave a target nucleic acid.

[353] In some embodiments, the repeat region may also be referred to as a "protein-binding segment." Typically, the repeat region is adjacent to the spacer region. For example, a guide RNA that interacts with an effector protein comprises a repeat region that is 5' of the spacer region.

V. Engineered Modifications

[354] Polypeptides (*e.g.*, effector proteins) and nucleic acids (*e.g.*, engineered guide nucleic acids) can be further modified as described herein. Examples are modifications that do not alter the primary sequence of the polypeptides or nucleic acids, including chemical derivatization of polypeptides (*e.g.*, acylation, acetylation, carboxylation, amidation, *etc.*). Also included are polypeptides that have a modified glycosylation pattern (*e.g.*, those made by: modifying the glycosylation patterns of a polypeptide during its synthesis and processing or in further processing steps; by exposing the polypeptide to enzymes which affect glycosylation, such as mammalian glycosylating or deglycosylating enzymes). Also embraced are polypeptides that have phosphorylated amino acid residues (*e.g.*, phosphotyrosine, phosphoserine, or phosphothreonine).

[355] Modifications disclosed herein can also include modification of described polypeptides and/or guide nucleic acids through any suitable method, such as molecular biological techniques and/or synthetic chemistry, to improve their resistance to proteolytic degradation, to change the target sequence specificity, to optimize solubility properties, to alter protein activity (*e.g.*, transcription modulatory activity, enzymatic activity, *etc.*) or to render them more suitable for their intended purpose (*e.g.*, *in vivo* administration, *in vitro* methods, or *ex vivo* applications). Analogs of such polypeptides include those containing residues other than naturally occurring L-amino acids, *e.g.* D-amino acids or non-naturally occurring synthetic amino acids. D-amino acids may be substituted for some or all of the amino acid residues. Modifications can also include modifications with non-naturally occurring unnatural amino acids. The particular sequence and the manner of preparation will be determined by convenience, economics, purity required, and the like.

[356] Modifications can further include the introduction of various groups to polypeptides and/or guide nucleic acids described herein. For example, groups can be introduced during synthesis or during expression of a polypeptide (*e.g.*, an effector protein), which allow for linking to other molecules or to a surface. Thus, *e.g.*, cysteines may be used to make thioethers, histidines for linking to a metal ion complex, carboxyl groups for forming amides or esters, amino groups for forming amides, and the like.

[357] Modifications can further include modification of nucleic acids described herein (*e.g.*, engineered guide nucleic acids) to provide the nucleic acid with a new or enhanced feature, such as improved stability. Such modifications of a nucleic acid include a base modification, a backbone modification, a sugar modification, or combinations thereof. In some embodiments, the modifications can be of one or more nucleotides, nucleosides, or nucleobases in a nucleic acid.

[358] In some embodiments, nucleic acids (*e.g.,* nucleic acids encoding effector proteins, engineered guide nucleic acids, or nucleic acids encoding engineered guide nucleic acids) described herein comprise one or more modifications comprising: 2'O-methyl modified nucleotides, 2' fluoro modified nucleotides; locked nucleic acid (LNA) modified nucleotides; peptide nucleic acid (PNA) modified nucleotides; nucleotides with phosphorothioate linkages; a 5' cap (*e.g.*, a 7-methylguanylate cap (m7G)), phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates, 5'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkyl phosphoramidates, phosphorodiamidates, thionophosphor amidates, thionoalkylphosphonates , thionoalkylphosphotriesters, selenophosphates and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein one or more internucleotide linkages is a 3' to 3', 5' to 5' or 2' to 2' linkage; phosphorothioate and/or heteroatom internucleoside linkages, such as $-CH_2-NH-O-CH_2-$, $-CH_2-N(CH_3)-O-CH_2-$ (known as a methylene (methylimino) or MMI backbone), $-CH_2-O-N(CH_3)-CH_2-$, $-CH_2-N(CH_3)-N(CH_3)-CH_2-$ and $-O-N(CH_3)-CH_2-CH_2-$ (wherein the native phosphodiester internucleotide linkage is represented as $-O-P(=O)(OH)-O-CH_2-$); morpholino linkages (formed in part from the sugar portion of a nucleoside);

morpholino backbones; phosphorodiamidate or other non-phosphodiester internucleoside linkages; siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; riboacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; other backbone modifications having mixed N, O, S and CH_2 component parts; and combinations thereof.

[359] In some embodiments, an engineered modification refers to a structural change of one or more nucleic acid residues of a nucleotide sequence or one or more amino acid residue of an amino acid sequence. The engineered modifications of a nucleotide sequence can include chemical modification of one or more nucleobases, or a chemical change to the phosphate backbone, a nucleotide, a nucleobase or a nucleoside. The engineered modifications can be made to an effector protein amino acid sequence or guide nucleic acid nucleotide sequence, or any sequence disclosed herein (*e.g.,* a nucleic acid encoding an effector protein or a nucleic acid that encodes a guide nucleic acid). Methods of modifying a nucleic acid or amino acid sequence are known. One of ordinary skill in the art will appreciate that the engineered modification(s) may be located at any position(s) of a nucleic acid such that the function of the nucleic acid, protein, composition or system is not substantially decreased. Nucleic acids provided herein can be prepared according to any available technique including, but not limited to chemical synthesis, enzymatic synthesis, which is generally termed in vitro-transcription, cloning, enzymatic, or chemical cleavage, etc. In some embodiments, the nucleic acids provided herein are not uniformly modified along the entire length of the molecule. Different nucleotide modifications and/or backbone structures can exist at various positions within the nucleic acid.

VI. Vectors and Multiplexed Expression Vectors

[360] Compositions, systems, and methods described herein comprise a vector or a use thereof. A vector can comprise one or more nucleic acids. In some embodiments, the nucleic acid of interest comprises one or more components of a composition or system described herein. In some embodiments, the nucleic acid of interest comprises a nucleotide sequence that encodes one or more components of the composition or system described herein. In some embodiments, one or more components comprises effector protein(s), fusion effector protein(s), fusion partner protein(s), guide nucleic acid(s), target nucleic acid(s), and donor nucleic acid(s). The vector may be part of a vector system, wherein a vector system comprises a library of vectors each encoding one or more component of a composition or system described herein. In some embodiments, components described herein (*e.g.,* an effector protein, a guide nucleic acid, and/or a target nucleic acid) are encoded by the same vector. In some embodiments, components described herein (*e.g.,* an effector protein, a guide nucleic acid, and/or a target nucleic acid) are each encoded by different vectors of the system.

[361] In some embodiments, a vector comprises a nucleotide sequence encoding one or more effector proteins as described herein. In some embodiments, the one or more effector proteins comprise at least

two effector proteins. In some embodiments, the at least two effector protein are the same. In some embodiments, the at least two effector proteins are different from each other. In some embodiments, the nucleotide sequence is operably linked to a promoter that is operable in a target cell, such as a eukaryotic cell. In some embodiments, the vector comprises the nucleotide sequence encoding 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 effector proteins.

[362] In some embodiments, a vector may encode one or more of any system components, including but not limited to effector proteins, guide nucleic acids, donor nucleic acids, and target nucleic acids as described herein. In some embodiments, a system component encoding a nucleotide sequence is operably linked to a promoter that is operable in a target cell, such as a eukaryotic cell. In some embodiments, a vector may encode 1, 2, 3, 4 or more of any system components. For example, a vector may encode two or more guide nucleic acids, wherein each guide nucleic acid comprises a different nucleotide sequence. A vector may encode an effector protein and a guide nucleic acid. A vector may encode an effector protein, a guide nucleic acid, and a donor nucleic acid.

[363] In some embodiments, a vector comprises one or more guide nucleic acids, or a nucleotide sequence encoding the one or more guide nucleic acids as described herein. In some embodiments, the one or more guide nucleic acids comprise at least two guide nucleic acids. In some embodiments, the at least two guide nucleic acids are the same. In some embodiments, the at least two guide nucleic acids are different from each other. In some embodiments, the guide nucleic acid or the nucleotide sequence encoding the guide nucleic acid is operably linked to a promoter that is operable in a target cell, such as a eukaryotic cell. In some embodiments, the vector comprises 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 guide nucleic acids. In some embodiments, the vector comprises a nucleotide sequence encoding 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 guide nucleic acids.

[364] In some embodiments, a vector comprises one or more donor nucleic acids as described herein. In some embodiments, the one or more donor nucleic acids comprise at least two donor nucleic acids. In some embodiments, the at least two donor nucleic acids are the same. In some embodiments, the at least two donor nucleic acids are different from each other. In some embodiments, the vector comprises 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 donor nucleic acids.

[365] In some embodiments, a fusion effector protein as described herein is inserted into a vector. In some embodiments, a vector may comprise or encode one or more regulatory elements. Regulatory elements may refer to transcriptional and translational control sequences, such as promoters, enhancers, polyadenylation signals, terminators, protein degradation signals, and the like, that provide for and/or regulate transcription of a non-coding sequence (*e.g.*, a guide nucleic acid) or a coding sequence (*e.g.*,

effector proteins, fusion proteins, and the like) and/or regulate translation of an encoded polypeptide. In some embodiments, the vector comprises one or more promoters, enhancers, ribosome binding sites, RNA splice sites, polyadenylation sites, a replication origin, and/or transcriptional terminator sequences. In some embodiments, a vector may comprise or encode for one or more additional elements, such as, for example, replication origins, antibiotic resistance (or a nucleic acid encoding the same), a tag (or a nucleic acid encoding the same), selectable markers, and the like.

[366] Vectors described herein can encode a promoter - a regulatory region on a nucleic acid, such as a DNA sequence, capable of initiating transcription of a downstream (3′ direction) coding or non-coding sequence. A promoter can be linked at its 3′ terminus to a nucleic acid, the expression or transcription of which is desired, and extends upstream (5′ direction) to include bases or elements necessary to initiate transcription or induce expression, which could be measured at a detectable level. A promoter can comprise a nucleotide sequence, referred to herein as a "promoter sequence". The promoter sequence can include a transcription initiation site, and one or more protein binding domains responsible for the binding of transcription machinery, such as RNA polymerase. When eukaryotic promoters are used, such promoters can contain "TATA" boxes and "CAT" boxes. Various promoters, including inducible promoters, may be used to drive expression, *i.e.*, transcriptional activation, of the nucleic acid of interest. Accordingly, in some embodiments, the nucleic acid of interest can be operably linked to a promoter.

[367] Promotors may be any suitable type of promoter envisioned for the compositions, systems, and methods described herein. Examples include constitutively active promoters (*e.g.,* CMV promoter), inducible promoters (*e.g.,* heat shock promoter, tetracycline-regulated promoter, steroid-regulated promoter, metal-regulated promoter, estrogen receptor-regulated promoter, etc.), spatially restricted and/or temporally restricted promoters (*e.g.,* a tissue specific promoter, a cell type specific promoter, etc.), etc. Suitable promoters include, but are not limited to: SV40 early promoter, mouse mammary tumor virus long terminal repeat (LTR) promoter; adenovirus major late promoter (Ad MLP); a herpes simplex virus (HSV) promoter, a cytomegalovirus (CMV) promoter such as the CMV immediate early promoter region (CMVIE), a rous sarcoma virus (RSV) promoter, a human U6 small nuclear promoter (U6), an enhanced U6 promoter, and a human Hl promoter (Hl). By transcriptional activation, it is intended that transcription will be increased above basal levels in the target cell by 2 fold, 5 fold, 10 fold, 50 fold, by 100 fold, 500 fold, or by 1000 fold, or more. In addition, vectors used for providing a nucleic acid that, when transcribed, produces a guide nucleic acid and/or a nucleic acid that encodes an effector protein to a cell may include nucleic acid sequences that encode for selectable markers in the target cells, so as to identify cells that have taken up the guide nucleic acid and/or the effector protein.

[368] In general, plasmids and vectors described herein comprise at least one promoter. In some embodiments, the promotors drive expression or transcription of one or more genome editing tools described herein. In some embodiments, the vector comprises a nucleotide sequence of a promoter. In some embodiments, the vector comprises two promoters. In some embodiments, the vector comprises

three promoters. In some embodiments, a length of the promoter is less than about 500, less than about 400, less than about 300, or less than about 200 linked nucleotides. In some embodiments, a length of the promoter is at least 100, at least 200, at least 300, at least 400, or at least 500 linked nucleotides. In some embodiments, the promoters are constitutive promoters. In other embodiments, the promoters are inducible promoters. In additional embodiments, the promoters are prokaryotic promoters (*e.g.*, drive expression of a gene in a prokaryotic cell). In some embodiments, the promoters are eukaryotic promoters, (*e.g.*, drive expression of a gene in a eukaryotic cell). Exemplary promoters include, but are not limited to, CMV, 7SK, EF1a, RPBSA, hPGK, SV40, PGK1, Ubc, human beta actin, CAG, TRE, UAS, Ac5, polyhedron, CaMKIIa, GAL1-10, H1, TEF1, GDS, ADH1, CaMV35S, Ubi, H1, U6, MSCV, MNDU3, and HSV TK promoter. In some embodiments, the promoter is CMV. In some embodiments, the promoter is EF1a. In some embodiments, the promoter is ubiquitin. In some embodiments, vectors are bicistronic or polycistronic vector (*e.g.*, having or involving two or more loci responsible for generating a protein) having an internal ribosome entry site (IRES) is for translation initiation in a cap-independent manner. In some embodiments, the promoter is an inducible promoter that only drives expression of its corresponding gene when a signal is present, *e.g.*, a hormone, a small molecule, a peptide. Non-limiting examples of inducible promoters are the T7 RNA polymerase promoter, the T3 RNA polymerase promoter, the Isopropyl-beta-D-thiogalactopyranoside (IPTG)-regulated promoter, a lactose induced promoter, a heat shock promoter, a tetracycline-regulated promoter (tetracycline-inducible or tetracycline-repressible), a steroid regulated promoter, a metal-regulated promoter, and an estrogen receptor-regulated promoter. In some embodiments, the promoter is an activation-inducible promoter, such as a CD69 promoter, as described further in Kulemzin et al., (2019), BMC Med Genomics, 12:44. In some embodiments, the promoter for expressing effector protein is a liver-specific promoter. In some embodiments, the liver-specific promoter comprises ApoE or TBG promoter sequence. In some embodiments, the promoter for expressing effector protein is a muscle-specific promoter. In some embodiments, the muscle-specific promoter comprises Ck8e, SPC5-12, or Desmin promoter sequence. In some embodiments, the promoter for expressing effector protein is a ubiquitous promoter. In some embodiments, the ubiquitous promoter comprises MND or CAG promoter sequence.

[369] In some embodiments, a vector described herein is a nucleic acid expression vector. In some embodiments, a vector described herein is a recombinant expression vector. In some embodiments, a vector described herein is a messenger RNA.

[370] In some examples, the delivery vector may be a eukaryotic vector, a prokaryotic vector (*e.g.*, a bacterial vector) a viral vector, or any combination thereof. In some embodiments, the delivery vehicle may be a non-viral vector. In some embodiments, the delivery vehicle may be a plasmid. In some embodiments, the plasmid comprises DNA. In some embodiments, the plasmid comprises RNA. In some examples, the plasmid comprises circular double-stranded DNA. In some examples, the plasmid may be linear. In some examples, the plasmid comprises one or more genes of interest and one or more

regulatory elements. In some examples, the plasmid comprises a bacterial backbone containing an origin of replication and an antibiotic resistance gene or other selectable marker for plasmid amplification in bacteria. In some examples, the plasmid may be a minicircle plasmid. In some examples, the plasmid contains one or more genes that provide a selective marker to induce a target cell to retain the plasmid. In some examples, the plasmid may be formulated for delivery through injection by a needle carrying syringe. In some examples, the plasmid may be formulated for delivery via electroporation. In some examples, the plasmids may be engineered through synthetic or other suitable means known in the art. For example, in some embodiments, the genetic elements may be assembled by restriction digest of the desired genetic sequence from a donor plasmid or organism to produce ends of the DNA which may then be readily ligated to another genetic sequence.

[371] In some embodiments, vectors comprise an enhancer. Enhancers are nucleotide sequences that have the effect of enhancing promoter activity. In some embodiments, enhancers augment transcription regardless of the orientation of their sequence. In some embodiments, enhancers activate transcription from a distance of several kilo basepairs. Furthermore, enhancers are located optionally upstream or downstream of a gene region to be transcribed, and/or located within the gene, to activate the transcription. Exemplary enhancers include, but are not limited to, WPRE; CMV enhancers; the R-U5′ segment in LTR of HTLV-I (Mol. Cell. Biol., Vol. 8(1), p. 466-472, 1988); SV40 enhancer; the intron sequence between exons 2 and 3 of rabbit β-globin (Proc. Natl. Acad. Sci. USA., Vol. 78(3), p. 1527-31, 1981); and the genome region of human growth hormone (J Immunol., Vol. 155(3), p. 1286-95, 1995).

Administration of a non-viral vector

[372] In some embodiments, the vector is a non-viral vector, and a physical method or a chemical method is employed for delivery into the somatic cell. Exemplary physical methods include electroporation, gene gun, sonoporation, magnetofection, or hydrodynamic delivery. Exemplary chemical methods include delivery of the recombinant polynucleotide via liposomes such as, cationic lipids or neutral lipids; dendrimers; nanoparticles; lipid nanoparticle (LNP); or cell-penetrating peptides.

[373] In some embodiments, a vector is administered as part of a method of nucleic acid detection, editing, and/or treatment as described herein. In some embodiments, a vector is administered in a single vehicle, such as a single expression vector. In some embodiments, at least two of the three components, a nucleic acid encoding one or more effector proteins, one or more donor nucleic acids, and one or more guide nucleic acids or a nucleic acid encoding the one or more guide nucleic acid, are provided in the single expression vector. In some embodiments, components, such as a guide nucleic acid and an effector protein, are encoded by the same vector. In some embodiments, an effector protein (or a nucleic acid encoding same) and/or an engineered guide nucleic acid (or a nucleic acid that, when transcribed, produces same) are not co-administered with donor nucleic acid in a single vehicle. In some

embodiments, an effector protein (or a nucleic acid encoding same), an engineered guide nucleic acid (or a nucleic acid that, when transcribed, produces same), and/or donor nucleic acid are administered in one or more or two or more vehicles, such as one or more, or two or more expression vectors.

[374] In some embodiments, a vector may be part of a vector system. In some embodiments, the vector system comprises a library of vectors each encoding one or more components of a composition or system described herein. In some embodiments, a vector system is administered as part of a method of nucleic acid detection, editing, and/or treatment as described herein, wherein at least two vectors are co-administered. In some embodiments, the at least two vectors comprise different components. In some embodiments, the at least two vectors comprise the same component having different sequences. In some embodiments, at least one of the three components, a nucleic acid encoding one or more effector proteins, one or more donor nucleic acids, and one or more guide nucleic acids or a nucleic acid encoding the one or more guide nucleic acids, or a variant thereof is provided in a different vector. In some embodiments, the nucleic acid encoding the effector protein, and a guide nucleic acid or a nucleic acid encoding the guide nucleic acid are provided in different vectors. In some embodiments, the donor nucleic acid is encoded by a different vector than the vector encoding the effector protein and the guide nucleic acid.

Lipid Particles and Non-viral Vectors

[375] In some embodiments, compositions and systems provided herein comprise a lipid particle. In some embodiments, a lipid particle is a lipid nanoparticle (LNP). In some embodiments, a lipid or a lipid nanoparticle can encapsulate an expression vector as described herein. LNPs are a non-viral delivery system for delivery of the composition and/or system components described herein. LNPs are particularly effective for delivery of nucleic acids. Beneficial properties of LNP include ease of manufacture, low cytotoxicity and immunogenicity, high efficiency of nucleic acid encapsulation and cell transfection, multi-dosing capabilities and flexibility of design (Kulkarni et al., (2018) Nucleic Acid Therapeutics, 28(3):146-157). In some embodiments, compositions and methods comprise a lipid, polymer, nanoparticle, or a combination thereof, or use thereof, to introduce one or more effector proteins, one or more guide nucleic acids, one or more donor nucleic acids, or any combinations thereof to a cell. Non-limiting examples of lipids and polymers are cationic polymers, cationic lipids, ionizable lipids, or bio-responsive polymers. In some embodiments, the ionizable lipids exploits chemical-physical properties of the endosomal environment (*e.g.*, pH) offering improved delivery of nucleic acids. In some embodiments, the ionizable lipids are neutral at physiological pH. In some embodiments, the ionizable lipids are protonated under acidic pH. In some embodiments, the bio-responsive polymer exploits chemical-physical properties of the endosomal environment (*e.g.*, pH) to preferentially release the genetic material in the intracellular space.

[376] In some embodiments, a LNP comprises an outer shell and an inner core. In some embodiments, the outer shell comprises lipids. In some embodiments, the lipids comprise modified lipids. In some

embodiments, the modified lipids comprise pegylated lipids. In some embodiments, the lipids comprise one or more of cationic lipids, anionic lipids, ionizable lipids, and non-ionic lipids. In some embodiments, the LNP comprises one or more of N1,N3,N5-tris(3-(didodecylamino)propyl)benzene-1,3,5-tricarboxamide (TT3), 2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE), 1-palmitoyl-2-oleoylsn-glycero-3-phosphoethanolamine (POPE), 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC), cholesterol (Chol), 1,2-dimyristoyl-sn-glycerol, and methoxypolyethylene glycol (DMG-PEChooo), derivatives, analogs, or variants thereof. In some embodiments, the LNP has a negative net overall charge prior to complexation with one or more of a guide nucleic acid, a nucleic acid encoding the one or more guide nucleic acid, a nucleic acid encoding the effector protein, and/or a donor nucleic acid. In some embodiments, the inner core is a hydrophobic core. In some embodiments, the one or more of a guide nucleic acid, the nucleic acid encoding the one or more guide nucleic acid, the nucleic acid encoding the effector protein, and/or the donor nucleic acid forms a complex with one or more of the cationic lipids and the ionizable lipids. In some embodiments, the nucleic acid encoding the effector protein or the nucleic acid encoding the guide nucleic acid is self-replicating.

[377] In some embodiments, a LNP comprises one or more of cationic lipids, ionizable lipids, and modified versions thereof. In some embodiments, the ionizable lipid comprises TT3 or a derivative thereof. Accordingly, in some embodiments, the LNP comprises one or more of TT3 and pegylated TT3. The publication WO2016187531 is hereby incorporated by reference in its entirety, which describes representative LNP formulations in **TABLE 2** and **TABLE 3**, and representative methods of delivering LNP formulations in Example 7.

[378] In some embodiments, a LNP comprises a lipid composition targeting to a specific organ. In some embodiments, the lipid composition comprises lipids having a specific alkyl chain length that controls accumulation of the LNP in the specific organ (*e.g.,* liver or spleen). In some embodiments, the lipid composition comprises a biomimetic lipid that controls accumulation of the LNP in the specific organ (*e.g.,* brain). In some embodiments, the lipid composition comprises lipid derivatives (*e.g.,* cholesterol derivatives) that controls accumulation of the LNP in a specific cell (*e.g.,* liver endothelial cells, Kupffer cells, hepatocytes).

[379] In some embodiments, the LNP described herein comprises nucleic acids (*e.g.*, DNA or RNA) encoding an effector protein described herein, an effector partner described herein, a fusion protein described herein, a guide nucleic acid described herein, or combinations thereof. In some embodiments, the LNP comprises an mRNA that produces an effector protein described herein, an effector partner described herein, or a fusion protein described herein when translated. In some embodiments, the LNP comprises chemically modified guide nucleic acids.

Delivery of Viral Vectors

[380] In some embodiments, a vector described herein comprises a viral vector. In some embodiments, the viral vector comprises a nucleic acid to be delivered into a host cell by a

recombinantly produced virus or viral particle. The nucleic acid may be single-stranded or double stranded, linear or circular, segmented or non-segmented. The nucleic acid may comprise DNA, RNA, or a combination thereof. In some embodiments, the vector is an adeno-associated viral vector. There are a variety of viral vectors that are associated with various types of viruses, including but not limited to retroviruses (*e.g.,* lentiviruses and γ-retroviruses), adenoviruses, arenaviruses, alphaviruses, adeno-associated viruses (AAVs), baculoviruses, vaccinia viruses, herpes simplex viruses and poxviruses. In some embodiments, the vector is an adeno-associated viral (AAV) vector. In some embodiments, the viral vector is a recombinant viral vector. The viral vector may be a retroviral vector. Retroviral vectors may include gamma-retroviral vectors such as vectors derived from the Moloney Murine Leukemia Virus (MoMLV, MMLV, MuLV, or MLV) or the Murine Stem cell Virus (MSCV) genome. Retroviral vectors may include lentiviral vectors such as those derived from the human immunodeficiency virus (HIV) genome. In some embodiments, the viral vector is a chimeric viral vector, comprising viral portions from two or more viruses. In some embodiments, the viral vector is a recombinant viral vector.

[381] In some embodiments, the viral vector corresponds to a virus of a specific serotype. In some embodiments, the viral vector is an AAV. The AAV may be any AAV known in the art. In some embodiments, the viral vector corresponds to a virus of a specific serotype. In some examples, the serotype is selected from an AAV1 serotype, an AAV2 serotype, AAV3 serotype, an AAV4 serotype, AAV5 serotype, an AAV6 serotype, AAV7 serotype, an AAV8 serotype, an AAV9 serotype, an AAV10 serotype, an AAV11 serotype, and an AAV12 serotype. In some embodiments the AAV vector is a recombinant vector, a hybrid AAV vector, a chimeric AAV vector, a self-complementary AAV (scAAV) vector, a single-stranded AAV or any combination thereof. scAAV genomes are generally known in the art and contain both DNA strands which can anneal together to form double-stranded DNA.

[382] In some embodiments, the AAV vector may be a chimeric AAV vector. In some embodiments, the chimeric AAV vector comprises an exogenous amino acid or an amino acid substitution, or capsid proteins from two or more serotypes. In some examples, a chimeric AAV vector may be genetically engineered to increase transduction efficiency, selectivity, or a combination thereof.

[383] In some embodiments, AAV vector described herein comprises two inverted terminal repeats (ITRs). According, in some embodiments, the viral vector provided herein comprises two inverted terminal repeats of AAV. A nucleotide sequence between the ITRs of an AAV vector provided herein comprises a nucleotide sequence encoding genome editing tools. In some embodiments, the genome editing tools comprise a nucleic acid encoding one or more effector proteins, a nucleic acid encoding one or more fusion proteins (*e.g.,* a nuclear localization signal (NLS), polyA tail), one or more guide nucleic acids, a nucleic acid encoding the one or more guide nucleic acids, respective promoter(s), one or more donor nucleic acid, or any combinations thereof. In some embodiments, viral vectors provided herein comprise at least one promotor or a combination of promoters driving expression or transcription of one or more genome editing tools described herein. In some embodiments, a coding region of the

AAV vector forms an intramolecular double-stranded DNA template thereby generating the AAV vector that is a self-complementary AAV (scAAV) vector. In some embodiments, the scAAV vector comprises the nucleotide sequence encoding genome editing tools that has a length of about 2 kb to about 3 kb. In some embodiments, the AAV vector provided herein is a self-inactivating AAV vector. In some embodiments, the AAV vector provided herein comprises a modification, such as an insertion, deletion, chemical alteration, or synthetic modification, relative to a wild-type AAV vector.

Producing AAV Delivery Vectors

[384] In some embodiments, methods of producing delivery vectors herein comprise packaging a nucleic acid encoding an effector protein and a guide nucleic acid, or a combination thereof, into an AAV vector. In some embodiments, methods of producing the delivery vector comprises, (a) contacting a cell with at least one nucleic acid encoding: (i) a guide nucleic acid; (ii) a Replication (Rep) gene; and (iii) a Capsid (Cap) gene that encodes an AAV capsid protein; (b) expressing the AAV capsid protein in the cell; (c) assembling an AAV particle; and (d) packaging a Cas effector encoding nucleic acid into the AAV particle, thereby generating an AAV delivery vector. In some embodiments, promoters, stuffer sequences, and any combination thereof may be packaged in the AAV vector. In some examples, the AAV vector can package 1, 2, 3, 4, or 5 guide nucleic acids or copies thereof. In some embodiments, the AAV vector comprises inverted terminal repeats, *e.g.*, a 5' inverted terminal repeat and a 3' inverted terminal repeat. In some embodiments, the AAV vector comprises a mutated inverted terminal repeat that lacks a terminal resolution site.

[385] In some embodiments, a hybrid AAV vector is produced by transcapsidation, *e.g.*, packaging an inverted terminal repeat (ITR) from a first serotype into a capsid of a second serotype, wherein the first and second serotypes may be not the same. In some examples, the Rep gene and ITR from a first AAV serotype (*e.g.*, AAV2) may be used in a capsid from a second AAV serotype (*e.g.*, AAV9), wherein the first and second AAV serotypes may be not the same. As a non-limiting example, a hybrid AAV serotype comprising the AAV2 ITRs and AAV9 capsid protein may be indicated AAV2/9. In some examples, the hybrid AAV delivery vector comprises an AAV2/1, AAV2/2, AAV 2/4, AAV2/5, AAV2/8, or AAV2/9 vector.

Producing AAV Particles

[386] In some embodiments, AAV particles described herein are recombinant AAV (rAAV). In some embodiments, rAAV particles are generated by transfecting AAV producing cells with an AAV-containing plasmid carrying the nucleotide sequence encoding the genome editing tools, a plasmid that carries viral encoding regions, i.e., Rep and Cap gene regions; and a plasmid that provides the helper genes such as E1A, E1B, E2A, E4ORF6 and VA. In some embodiments, the AAV producing cells are mammalian cells. In some embodiments, host cells for rAAV viral particle production are mammalian cells. In some embodiments, a mammalian cell for rAAV viral particle production is a COS cell, a

HEK293T cell, a HeLa cell, a KB cell, a variant thereof, or a combination thereof. In some embodiments, rAAV virus particles can be produced in the mammalian cell culture system by providing the rAAV plasmid to the mammalian cell. In some embodiments, producing rAAV virus particles in a mammalian cell comprises transfecting vectors that express the rep protein, the capsid protein, and the gene-of-interest expression construct flanked by the ITR sequence on the 5' and 3' ends. Methods of such processes are provided in, for example, Naso et al., BioDrugs, 2017 Aug;31(4):317-334 and Benskey et al., (2019), Methods Mol Biol., 1937:3-26, each of which is incorporated by reference in their entireties.

[387] In some embodiments, rAAV is produced in a non-mammalian cell. In some embodiments, rAAV is produced in an insect cell. In some embodiments, the insect cell for producing rAAV viral particles comprises a Sf9 cell. In some embodiments, production of rAAV virus particles in insect cells may comprise baculovirus. In some embodiments, production of rAAV virus particles in insect cells may comprise infecting the insect cells with three recombinant baculoviruses, one carrying the cap gene, one carrying the rep gene, and one carrying the gene-of-interest expression construct enclosed by an ITR on both the 5' and 3' end. In some embodiments, rAAV virus particles are produced by the One Bac system. In some embodiments, rAAV virus particles can be produced by the Two Bac system. In some embodiments, in the Two Bac system, the rep gene and the cap gene of the AAV is integrated into one baculovirus virus genome, and the ITR sequence and the gene-of-interest expression construct is integrated into another baculovirus virus genome. In some embodiments, in the One Bac system, an insect cell line that expresses both the rep protein and the capsid protein is established and infected with a baculovirus virus integrated with the ITR sequence and the gene-of-interest expression construct. Details of such processes are provided in, for example, Smith et. al., (1983), Mol. Cell. Biol., 3(12):2156-65; Urabe et al., (2002), Hum. Gene. Ther., 1;13(16):1935-43; and Benskey et al., (2019), Methods Mol Biol., 1937:3-26, each of which is incorporated by reference in its entirety.

VII. Target Nucleic Acids

[388] Disclosed herein are compositions, systems and methods for editing a target nucleic acid. In some embodiments, the target nucleic acid is a single stranded nucleic acid. Alternatively, or in combination, the target nucleic acid is a double stranded nucleic acid and is prepared into single stranded nucleic acids before or upon contacting the reagents. In some embodiments, the target nucleic acid is a double stranded nucleic acid. In some embodiments, the double stranded nucleic acid is DNA. The target nucleic acid may be an RNA. The target nucleic acids include but are not limited to mRNA, rRNA, tRNA, non-coding RNA, long non-coding RNA, and microRNA (miRNA). In some embodiments, the target nucleic acid is complementary DNA (cDNA) synthesized from a single-stranded RNA template in a reaction catalyzed by a reverse transcriptase. In some embodiments, the target nucleic acid is single-stranded RNA (ssRNA) or mRNA. In some embodiments, the target nucleic

acid is mRNA. In some embodiments, the target nucleic acid is from a virus, a parasite, or a bacterium described herein.

[389] In some embodiments, an effector protein or a multimeric complex thereof recognizes a PAM on a target nucleic acid. In some embodiments, multiple effector proteins of the multimeric complex recognize a PAM on a target nucleic acid. In some embodiments, only one effector protein of the multimeric complex recognizes a PAM on a target nucleic acid. In some embodiments, the PAM is 3' to the spacer region of the guide nucleic acid (*e.g.*, a crRNA or sgRNA) described herein. In some embodiments, the PAM is directly 3' to the spacer region of the guide nucleic acid (*e.g.*, a crRNA or sgRNA) described herein. In some embodiments, the PAM sequence comprises a nucleotide sequence listed in **TABLE 3**.

[390] An effector protein of the present disclosure, a dimer thereof, or a multimeric complex thereof may cleave or nick a target nucleic acid within or near a protospacer adjacent motif (PAM) sequence of the target nucleic acid. In some embodiments, cleavage occurs within 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 nucleosides of a 5' or 3' terminus of a PAM sequence. A target nucleic acid may comprise a PAM sequence adjacent to a sequence that is complementary to a guide nucleic acid spacer region. In some embodiments, the PAM sequence is read 5' to 3' as set forth in **TABLE 3**.

[391] In some embodiments, the effector protein comprises an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identical to the amino acid sequence of **TABLE 1A** and **TABLE 1B**, and the target nucleic acid comprises a PAM sequence of any one of the nucleotide sequences as set forth in **TABLE 3**.

[392] In some embodiments, a target nucleic acid comprising a target sequence comprises a PAM sequence. In some embodiments, the PAM sequence is 3' to the target sequence. In some embodiments, the PAM sequence is directly 3' to the target sequence. In some embodiments, the PAM sequence 5' to the target sequence. In some embodiments, the PAM sequence is directly 5' to the target sequence. In some embodiments, the target nucleic acid as described in the methods herein does not initially comprise a PAM sequence. However, any target nucleic acid of interest may be generated using the methods described herein to comprise a PAM sequence, and thus be a PAM target nucleic acid. A PAM target nucleic acid, as used herein, refers to a target nucleic acid that has been amplified to insert a PAM sequence that is recognized by an effector system described herein.

[393] In some embodiments, the target nucleic acid comprises 5 to 100, 5 to 90, 5 to 80, 5 to 70, 5 to 60, 5 to 50, 5 to 40, 5 to 30, 5 to 25, 5 to 20, 5 to 15, or 5 to 10 linked nucleosides. In some embodiments, the target nucleic acid comprises 10 to 90, 20 to 80, 30 to 70, or 40 to 60 linked nucleosides. In some embodiments, the target nucleic acid comprises 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, 45, 50, 60, 70, 80, 90, or 100 linked nucleosides. In some embodiments, the target nucleic acid comprises at least 10, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, or at least 100 linked nucleosides.

[394] In some embodiments, a target nucleic acid comprises a portion or a specific region of a nucleic acid from a genomic locus from a gene described herein. Non-limiting examples of genes are recited in **TABLE 10**. Nucleic acid sequences of target nucleic acids and/or corresponding genes are readily available in public databases as known and used in the art. In some embodiments, the target nucleic acid is selected from **TABLE 10**. In some embodiments, the target nucleic acid comprises one or more target sequences. In some embodiments, the one or more target sequence is within any one of the target nucleic acids set forth in **TABLE 10**. In some embodiments, the target nucleic acid is a eukaryotic gene. In some embodiments, the target nucleic acid is a mammalian gene. In some embodiments, the target nucleic acid is a human gene. In some embodiments, the target nucleic acid sequence comprises a nucleotide sequence within a human safe harbor locus.

[395] In some embodiments, the human safe harbor locus is located in human chromosome 1, human chromosome 3, human chromosome 4, human chromosome 6, human chromosome 10, human chromosome 11, human chromosome 12, human chromosome 14, human chromosome 17, human chromosome 18, or human chromosome 19. In some embodiments, the human safe harbor locus comprises any one of: exon 1-2 of Angptl3 (human chromosome 1), exon 1-2 of AAVS1 (PPP1R12C) (human chromosome 19), exon 1-2 of ALB (human chromosome 4), exon 1-2 of ApoC3 (human chromosome 11), exon 1-2 of ASGR2 (human chromosome 17), exon 1-2 of CCR5 (human chromosome 3), exon 1-2 of FIX (F9) (human chromosome 10), exon 1-2 of Gys2 (human chromosome 12), exon 1-2 of HGD (human chromosome 3), exon 1-2 of Lp(a) (human chromosome 6), exon 1-2 of Pcsk9 (human chromosome 1), exon 1-2 of Serpina1 (human chromosome 14), exon 1-2 of TF (human chromosome 3), and exon 1-2 of TTR (human chromosome 18). In some embodiments, the human safe harbor locus comprises at least one sequence in AAVS1, CCR5, human ortholog sequence of the mouse Rosa26, or human albumin (also referred to as human serum albumin or HSA or ALB). In some instances, the human safe harbor locus is located in human chromosome 2 or human chromosome 4. In some embodiments, the targeting the human safe harbor loci as described herein for incorporation of a transgene as described herein results in stable expression of the functional human protein and wherein the incorporation does not result in insertional oncogenesis.

[396] In some embodiments, the target nucleic acid is within albumin gene. In some embodiments, the target nucleic acid is within human albumin gene. The human albumin gene is located at chromosome 4q11-q13. In some embodiments, the target nucleic acid is within intron 1 of albumin gene. In some embodiments, the target nucleic acid is within intron 1 of human albumin gene.

[397] **TABLE 10** provides exemplary sequences for use with the compositions and methods of the disclosure. In some embodiments, the target sequence comprises a nucleotide sequence that is at least 70%, at least 80%, at least 90%, at least 92%, at least 95%, at least 97%, at least 99%, or 100% identical to the nucleotide sequence recited in **TABLE 10**.

[398] In some embodiments, at least a portion of the target nucleic acid that a guide nucleic acid binds is within a nucleotide sequence about 5 or more, about 10 or more, about 15 or more, about 20 or more,

about 25 or more, about 30 or more, about 35 or more, about 40 or more, about 45 or more, about 50 or more, about 55 or more, about 60 or more, about 65 or more, about 70 or more, about 75 or more, about 80 or more, about 85 or more, about 90 or more, about 95 or more, about 100 or more, about 105 or more, about 110 or more, about 115 or more, about 120 or more, about 125 or more, about 130 or more, about 135 or more, about 140 or more, about 145 or more, or about 150 or more nucleotides adjacent to: the start of a targeted intron, the end of a targeted intron, or both.

[399] In some embodiments, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid comprises a spacer sequence, wherein the spacer sequence hybridizes to a target sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to **SEQ ID NO: 266**. In some embodiments, the spacer sequence is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of the nucleotide sequences recited in **TABLE 4.**

[400] In some embodiments, the guide nucleic acid comprises a spacer sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of **SEQ ID NO: 301-604**, wherein an effector protein has an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of **SEQ ID NO: 6-12, SEQ ID NO: 228-231**, **47**, and **13**, and wherein the spacer sequence hybridizes to a target sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to **SEQ ID NO: 266**. In some embodiments, the guide nucleic acid comprises the spacer sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of **SEQ ID NO: 301-510**, wherein the effector protein has an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of **SEQ ID NO: 6-12,** and **SEQ ID NO: 228-230**. In some embodiments, the guide nucleic acid comprises the spacer sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of **SEQ ID NO: 511-535**, wherein the effector protein has an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 13**. In some embodiments, the guide nucleic acid comprises the spacer sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of **SEQ ID NO: 536-575**, wherein the effector protein has an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 47**. In some embodiments, the guide nucleic acid comprises the spacer sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of **SEQ ID NO: 576-604**, wherein the effector protein has an amino acid sequence that is at least 65%, at least 70%, at

least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to **SEQ ID NO: 231**.

[401] In some embodiments, the guide nucleic acid or a nucleic acid encoding the guide nucleic acid comprises a spacer sequence and/or a repeat sequence, wherein the spacer sequence is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of the nucleotide sequences recited in **TABLE 4**, wherein the repeat sequence is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of the nucleotide sequences recited in **TABLE 6**, and wherein the spacer sequence hybridizes to a target sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to **SEQ ID NO: 266**.

[402] In some embodiments, compositions disclosed herein comprises an effector protein comprising an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% identical to any one of **SEQ ID NO: 6-12,** and **228-230**; a crRNA comprising a repeat sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to **SEQ ID NO: 252 or 1848**, and a spacer sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of **SEQ ID NO: 301-510**, wherein the spacer sequence hybridizes to a target sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to **SEQ ID NO: 266**.

[403] In some embodiments, compositions disclosed herein comprises an effector protein comprising an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% identical to **SEQ ID NO: 13**; a crRNA comprising a repeat sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to **SEQ ID NO: 253**, and a spacer sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of **SEQ ID NO: 511-535**, wherein the spacer sequence hybridizes to a target sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to **SEQ ID NO: 266**.

[404] In some embodiments, compositions disclosed herein comprises an effector protein comprising an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% identical to **SEQ ID NO: 47**; a crRNA comprising a repeat sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to **SEQ ID NO: 254**, and a spacer sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of **SEQ ID NO: 536-575**, wherein the spacer sequence hybridizes to a target sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to **SEQ ID NO: 266**.

[405] In some embodiments, compositions disclosed herein comprises an effector protein comprising an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at

least 95%, at least 99%, or 100% identical to **SEQ ID NO: 231**; a crRNA comprising a repeat sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to **SEQ ID NO: 255** or **1789**, and a spacer sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of **SEQ ID NO: 576-604**, wherein the spacer sequence hybridizes to a target sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to **SEQ ID NO: 266**.

[406] In some embodiments, the guide nucleic acid or a nucleic acid encoding the guide nucleic acid comprises a spacer sequence and/or a handle sequence, wherein the spacer sequence is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of the nucleotide sequences recited in **TABLE 4**, wherein the handle sequence is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of the nucleotide sequences recited in **TABLE 7**, and wherein the spacer sequence hybridizes to a target sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to **SEQ ID NO: 266**. In some embodiments, the handle sequence comprises one or more of an intermediary sequence, a repeat sequence, and a linker.

[407] In some embodiments, compositions disclosed herein comprises an effector protein comprising an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% identical to **SEQ ID NO: 13**; a guide nucleic acid comprising a handle sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of **SEQ ID NO: 259-261**, and a spacer sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of **SEQ ID NO: 511-535**, wherein the spacer sequence hybridizes to a target sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to **SEQ ID NO: 266**.

[408] In some embodiments, compositions disclosed herein comprises an effector protein comprising an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% identical to **SEQ ID NO: 231**; a guide nucleic acid comprising a handle sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of **SEQ ID NO: 262-264**, and a spacer sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of **SEQ ID NO: 576-604**, wherein the spacer sequence hybridizes to a target sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to **SEQ ID NO: 266**.

[409] In some embodiments, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid comprises a spacer sequence, wherein the spacer sequence hybridizes to a target sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to **SEQ ID NO: 266**. In some embodiments, the spacer sequence is at least 70%, at least 75%, at least 80%, at least

85%, at least 90%, at least 95%, or 100% identical to any one of the nucleotide sequences recited in **TABLE 4.**

[410] Further description of editing or detecting a target nucleic acid in the foregoing genes can be found in more detail in Kim et al., "*Enhancement of target specificity of CRISPR-Cas12a by using a chimeric DNA-RNA guide*", Nucleic Acids Res. 2020 Sep 4;48(15):8601-8616; Wang et al., "*Specificity profiling of CRISPR system reveals greatly enhanced off-target gene editing*", Scientific Reports volume 10, Article number: 2269 (2020); Tuladhar et al., "*CRISPR-Cas9-based mutagenesis frequently provokes on-target mRNA misregulation*", Nature Communications volume 10, Article number: 4056 (2019); Dong et al., "*Genome-Wide Off-Target Analysis in CRISPR-Cas9 Modified Mice and Their Offspring*", G3, Volume 9, Issue 11, 1 November 2019, Pages 3645–3651; Winter et al., "*Genome-wide CRISPR screen reveals novel host factors required for Staphylococcus aureus α-hemolysin-mediated toxicity*", Scientific Reports volume 6, Article number: 24242 (2016); and Ma et al., "*A CRISPR-Based Screen Identifies Genes Essential for West-Nile-Virus-Induced Cell Death*", Cell Rep. 2015 Jul 28;12(4):673-83, which are hereby incorporated by reference in their entirety.

[411] In some embodiments, the target sequence comprises at least 15, at least 16, at least 17, at least 18, at least 19, or at least 20 contiguous nucleotides of any one of the nucleotide sequences recited in **TABLE 10**, a complement thereof, or a reverse complement thereof. In some embodiments, the target sequence comprises a nucleotide sequence that is at least 70%, at least 80%, at least 90%, at least 92%, at least 95%, at least 97%, at least 99%, or 100% identical to the nucleotide sequence recited in **TABLE 10**. In some embodiments, the target nucleic acid comprises a target sequence, wherein the target sequence comprises at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, or at least 25 contiguous nucleotides of any one of the nucleotide sequences recited in **TABLE 10**. In some embodiments, a target nucleic acid comprises 5 to 100, 5 to 90, 5 to 80, 5 to 70, 5 to 60, 5 to 50, 5 to 40, 5 to 30, 5 to 25, 5 to 20, 5 to 15, or 5 to 10 linked nucleotides of any one of the nucleotide sequences recited in **TABLE 10**. In some embodiments, the target nucleic acid comprises 10 to 90, 20 to 80, 30 to 70, or 40 to 60 linked nucleotides of any one of the nucleotide sequences recited in **TABLE 10**. In some embodiments, the target nucleic acid comprises 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, 45, 50, 60, 70, 80, 90, or 100 linked nucleotides of any one of the nucleotide sequence recited in **TABLE 10**. In some embodiments, the target nucleic acid comprises at least 10, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, or at least 100 linked nucleotides of any one of the nucleotide sequences recited in **TABLE 10**.

[412] In some embodiments, compositions, systems, and methods described herein comprise a modified target nucleic acid which can describe a target nucleic acid wherein the target nucleic acid has undergone a modification, for example, after contact with an effector protein. In some embodiments, the modification is an alteration in the sequence of the target nucleic acid. In some embodiments, the

modified target nucleic acid comprises an insertion, deletion, or replacement of one or more nucleotides compared to the unmodified target nucleic acid. In some embodiments, the modification is a mutation.

[413] In some embodiments, the target nucleic acid is in a cell. In some embodiments, the cell is a hepatocyte. In some embodiments, the cell is a human cell. In some embodiments, the human cell is a hepatocyte. stem cell, progenitor cell, induced pluripotent stem cell (iPSC), or a cell derived from an iPSC cell.

Donor Nucleic Acids

[414] In some embodiments, a donor nucleic acid comprises a nucleic acid that is incorporated into a target nucleic acid or target sequence.

[415] In some embodiments, a donor nucleic acid comprises a transgene. In some embodiments, the transgene comprises a nucleotide sequence that is inserted into a cell for expression of said nucleotide sequence in the cell. In some embodiments, the transgene comprises (1) a nucleotide sequence that is not naturally found in the cell (*e.g.,* a heterologous nucleotide sequence); (2) a nucleotide sequence that is a mutant form of a nucleotide sequence naturally found in the cell into which it has been introduced; (3) a nucleotide sequence that serves to add additional copies of the same (*e.g.,* exogenous or homologous) or a similar nucleotide sequence naturally occurring in the cell into which it has been introduced; or (4) a silent naturally occurring or homologous nucleotide sequence whose expression is induced in the cell into which it has been introduced. A donor nucleic acid can comprise a transgene. The cell in which transgenes expression occur can be a target cell, such as a host cell.

[416] In some embodiments, transgenes described herein can be inserted or integrated into a target nucleic acid. In some embodiments, the donor nucleic acid comprises a human gene. In some embodiments, the donor nucleic acid comprises a gene encoding a human functional protein. Functional human proteins as described herein include proteins that are encoded by transgenes incorporated in donor nucleic acids described herein. Specific examples of functional human proteins include wildtype and engineered versions of proteins that maybe deficient, under-expressed or aberrantly expressed in certain human subjects. In some instances, these subjects suffer from disorders that result in deficiency, or aberrant expression of these proteins. Compositions and methods provided herein are useful to treat conditions characterized by such deficiency, or aberrant expression. In some embodiments, a compositions and systems comprise a donor nucleic acid encoding a functional protein that is poorly expressed in subjects with genetic disorders such as monogenic disorders. In some embodiments, a functional protein refers to protein that retains at least some if not all activity relative to the wildtype protein. A functional protein can also include a protein having enhanced activity relative to the wildtype protein. Assays are known and available for detecting and quantifying protein activity, *e.g.,* colorimetric and fluorescent assays. In some embodiments, a functional protein is a wildtype protein. In some embodiments, a functional protein is a functional portion of a wildtype protein. By way of non-limiting example, functional human proteins that are poorly expressed in subjects with genetic disorders include

CFTR, DMD, A1AT, GAA, FXN, F8, F9, SOD1, C9, HTT, MECP2, SMN1, TARDBP, FUS, RHO, and USH2A. Exemplary amino acid sequences of functional human proteins are provided in **TABLE 11**.

[417] In some embodiments, are methods of administering to the subject a donor nucleic acid described herein, wherein the subject comprises a mutation that results in a reduction or loss of a functional protein. In some embodiments, a mutation comprises a point mutation or single nucleotide polymorphism (SNP), a chromosomal mutation, a copy number mutation, or any combination thereof. A point mutation optionally comprises a substitution, insertion, or deletion. In some embodiments, a mutation comprises a chromosomal mutation. A chromosomal mutation can comprise an inversion, a deletion, a duplication, or a translocation. In some embodiments, a mutation comprises a copy number variation. A copy number variation can comprise a gene amplification or an expanding trinucleotide repeat. The mutation may be located in a non-coding region or a coding region of a gene.

[418] A donor nucleic acid may be inserted at cleavage site within the target nucleic acid, wherein the cleavage site is generated by an effector protein or fusion protein described herein. In some embodiments, the donor nucleic acid encodes amino acid sequence of a functional human protein. In some embodiments, the functional human protein has 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 180%, at least 200%, at least 300%, at least 400% enzymatic activity compared to wildtype. In some embodiments, the functional human protein comprises wildtype. In some embodiments, the wildtype protein comprises human wildtype protein sequence. In some embodiments, the human protein comprises an amino acid sequence recited in **TABLE 11**. In some embodiments, the donor nucleic acid encoding amino acid sequence that is at least 70%, at least 80%, at least 90%, at least 92%, at least 95%, at least 97%, at least 99%, or 100% identical to any one of the amino acid sequences recited in **TABLE 11**. In some embodiments, methods comprise contacting a target nucleic acid with an effector protein comprising an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**, thereby introducing a single-stranded break in the target nucleic acid; contacting the target nucleic acid with a second effector protein, optionally comprising an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99% identical to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**, to generate a second cleavage site in the target nucleic acid, ligating the regions flanking the first and second cleavage site, optionally through NHEJ or single-strand annealing, thereby resulting in the excision of a portion of the target nucleic acid between the first and second cleavage sites from the target nucleic acid; and contacting the target nucleic acid with a donor nucleic acid for homologous recombination, optionally via HDR or NHEJ, thereby introducing a new sequence into the target nucleic acid (*e.g.*, at a cleavage site or in between two cleavage sites).

VIII. Compositions

[419] Disclosed herein are compositions comprising one or more effector proteins described herein or nucleic acids encoding the one or more effector proteins, one or more guide nucleic acids described herein or nucleic acids encoding the one or more guide nucleic acids described herein, or combinations thereof. In some embodiments, one or more of a repeat sequence, a handle sequence, and intermediary sequence of the one or more guide nucleic acids are capable of interacting with the one or more of the effector proteins. In some embodiments, spacer sequences of the one or more guide nucleic acids hybridizes with a target sequence of a target nucleic acid. In some embodiments, the compositions comprise one or more donor nucleic acids described herein. In some embodiments, the compositions are capable of editing a target nucleic acid in a cell or a subject. In some embodiments, the compositions are capable of editing a target nucleic acid or the expression thereof in a cell, in a tissue, in an organ, in vitro, in vivo, or ex vivo. In some embodiments, the compositions are capable of editing a target nucleic acid in a sample comprising the target nucleic.

[420] Disclosed herein are compositions comprising: an effector protein, or a nucleic acid encoding the effector protein; a guide nucleic acid, or a nucleic acid encoding the guide nucleic acid; and a donor nucleic acid. In some embodiments, the effector protein comprises an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the guide nucleic acid is a crRNA. In some embodiments, the crRNA comprises a spacer sequence and a repeat sequence. In some embodiments, the guide nucleic acid is a single guide RNA (sgRNA). In some embodiments, the sgRNA comprises a spacer sequence and a handle sequence. In some embodiments, the handle sequence comprises one or more of an intermediary sequence, a repeat sequence, and a linker. In some embodiments, the guide nucleic acid comprises a spacer sequence that hybridizes to a target sequence in a target nucleic acid. In some embodiments, the target sequence comprises a nucleotide sequence within a human safe harbor locus. In some embodiments, the target sequence comprises any one of the nucleotide sequences recited in **TABLE 10**. In some embodiments, the donor nucleic acid encodes a transgene that comprises a functional human protein that is expressed upon incorporation into the human safe harbor locus.

[421] In some embodiments, compositions described herein comprise plasmids described herein, viral vectors described herein, non-viral vectors described herein, or combinations thereof. In some embodiments, compositions described herein comprise the viral vectors. In some embodiments, compositions described herein comprise an AAV. In some embodiments, compositions described herein comprise liposomes (*e.g.,* cationic lipids or neutral lipids), dendrimers, lipid nanoparticle (LNP), or cell-penetrating peptides. In some embodiments, compositions described herein comprise an LNP.

Pharmaceutical Compositions and Modes of Administration

[422] Disclosed herein, in some aspects, are pharmaceutical compositions for modifying a target nucleic acid in a cell or a subject, comprising any one of the effector proteins, engineered effector proteins, fusion effector proteins, or guide nucleic acids as described herein and any combination thereof. Also disclosed herein, in some aspects, are pharmaceutical compositions comprising a nucleic acid encoding any one of the effector proteins, engineered effector proteins, fusion effector proteins, guide nucleic acids, or donor nucleic acid as described herein and any combination thereof. In some embodiments, pharmaceutical compositions comprise a plurality of guide nucleic acids. Pharmaceutical compositions may be used to modify a target nucleic acid or the expression thereof in a cell in vitro, in vivo or ex vivo.

[423] In some embodiments, pharmaceutical compositions comprise one or more nucleic acids encoding an effector protein, fusion effector protein, fusion partner, a guide nucleic acid, or a combination thereof; and a pharmaceutically acceptable carrier or diluent. The effector protein, fusion effector protein, fusion partner protein, or combination thereof may be any one of those described herein. The one or more nucleic acids may comprise a plasmid. The one or more nucleic acids may comprise a nucleic acid expression vector. The one or more nucleic acids may comprise a viral vector. In some embodiments, the viral vector is a lentiviral vector. In some embodiments, the vector is an adeno-associated viral (AAV) vector. In some embodiments, compositions, including pharmaceutical compositions, comprise a viral vector encoding a fusion effector protein and a guide nucleic acid, wherein at least a portion of the guide nucleic acid binds to the effector protein of the fusion effector protein.

[424] In some embodiments, pharmaceutical compositions comprise a virus comprising a viral vector encoding a fusion effector protein, an effector protein, a fusion partner, a guide nucleic acid, or a combination thereof; and a pharmaceutically acceptable carrier or diluent. The virus may be a lentivirus. The virus may be an adenovirus. The virus may be a non-replicating virus. The virus may be an adeno-associated virus (AAV).

[425] Pharmaceutical compositions described herein may comprise a salt. In some embodiments, the salt is a sodium salt. In some embodiments, the salt is a potassium salt. In some embodiments, the salt is a magnesium salt. In some embodiments, the salt is NaCl. In some embodiments, the salt is KNO3. In some embodiments, the salt is Mg^{2+} SO_4^{2-}.

[426] Non-limiting examples of pharmaceutically acceptable carriers and diluents suitable for the pharmaceutical compositions disclosed herein include buffers (*e.g.*, neutral buffered saline, phosphate buffered saline); carbohydrates (*e.g.*, glucose, mannose, sucrose, dextran, mannitol); polypeptides or amino acids (*e.g.*, glycine); antioxidants; chelating agents (*e.g.*, EDTA, glutathione); adjuvants (*e.g.*, aluminum hydroxide); surfactants (Polysorbate 80, Polysorbate 20, or Pluronic F68); glycerol; sorbitol; mannitol; polyethyleneglycol; and preservatives.

[427] In some embodiments, pharmaceutical compositions are in the form of a solution (*e.g.*, a liquid). The solution may be formulated for injection, *e.g.*, intravenous or subcutaneous injection. In some

embodiments, the pH of the solution is about 7, about 7.1, about 7.2, about 7.3, about 7.4, about 7.5, about 7.6, about 7.7, about 7.8, about 7.9, about 8, about 8.1, about 8.2, about 8.3, about 8.4, about 8.5, about 8.6, about 8.7, about 8.8, about 8.9, or about 9. In some embodiments, the pH is 7 to 7.5, 7.5 to 8, 8 to 8.5, 8.5 to 9, or 7 to 8.5. In some embodiments, the pH of the solution is less than 7. In some embodiments, the pH is greater than 7.

[428] In some embodiments, pharmaceutical compositions comprise an: effector protein, fusion effector protein, fusion partner, a guide nucleic acid, or a combination thereof; and a pharmaceutically acceptable carrier or diluent. In some embodiments, pharmaceutical compositions comprise one or more nucleic acids encoding an: effector protein, fusion effector protein, fusion partner, a guide nucleic acid, or a combination thereof; and a pharmaceutically acceptable carrier or diluent. In some embodiments, guide nucleic acid can be a plurality of guide nucleic acids. In some embodiments, the effector protein comprises an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the amino acid sequences of **TABLE 1A** and **TABLE 1B**.

[429] In some embodiments, pharmaceutical compositions comprise an: effector protein, fusion effector protein, fusion partner, a guide nucleic acid, or a combination thereof; and a pharmaceutically acceptable carrier or diluent. In some embodiments, pharmaceutical compositions comprise one or more nucleic acids encoding an: effector protein, fusion effector protein, fusion partner, a guide nucleic acid, or a combination thereof; and a pharmaceutically acceptable carrier or diluent. In some embodiments, guide nucleic acid can be a plurality of guide nucleic acids. In some embodiments, the effector protein comprises an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to the amino acid sequence recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the nucleotide sequence of the gRNA is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the gRNA sequences recited in **TABLE 4, TABLE 5, TABLE 6, TABLE 7, TABLE 8,** and **TABLE 9**. In some embodiments, the nucleotide sequence of the gRNA is at least 70%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to **SEQ ID NO: 1825, 1828, 1846** or **1847**.

[430] In combination with a pharmaceutically acceptable carrier or diluent, each row in **TABLE 13**, **TABLE 14**, **TABLE 15**, **TABLE 16**, **TABLE 17**, **TABLE 19**, **TABLE 20**, and **TABLE 21** can represent an exemplary pharmaceutical composition comprising an effector protein as set forth in **TABLE 1A** and **TABLE 1B** recognizing any one of the PAM sequences recited in **TABLE 3** and a guide nucleic acid wherein the guide nucleic acid is a gRNA. In some embodiments, the guide nucleic acid comprises a nucleotide sequence of any one of the gRNA sequences recited in **TABLE 4, TABLE 5, TABLE 6, TABLE 7, TABLE 8,** and **TABLE 9**. In some embodiments, the guide nucleic acid comprises a nucleotide sequence of **SEQ ID NO: 1825, 1828, 1846** or **1847**. In some embodiments, the nucleotide sequence of the gRNA is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%,

at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the gRNA sequences recited in **TABLE 4, TABLE 5, TABLE 6, TABLE 7, TABLE 8,** and **TABLE 9**. In some embodiments, the nucleotide sequence of the gRNA is at least 70%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to **SEQ ID NO: 1825, 1828, 1846** or **1847**.

IX. Systems

[431] Disclosed herein, in some aspects, are systems for modifying, or editing a target nucleic acid, comprising the effector proteins or nucleic acids encoding the effector proteins described herein, or a multimeric complex thereof. Systems may be used to modify or edit a target nucleic acid. Systems may be used to insert a donor nucleic acid into a target nucleic acid. In some embodiments, systems comprise an effector protein or a nucleic acid encoding the effector protein described herein, a guide nucleic acid or a nucleic acid encoding the guide nucleic acid a reagent described herein, a donor nucleic acid described herein, support medium, or a combination thereof. In some embodiments, the effector protein comprises an effector protein, or a fusion protein thereof, described herein. In some embodiments, effector proteins comprise an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of the amino acid sequences of **TABLE 1A** and **TABLE 1B**. In some embodiments, the amino acid sequence of the effector protein is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of the amino acid sequences of **TABLE 1A** and **TABLE 1B**. In some embodiments, the guide nucleic acid comprises at least one nucleotide sequence selected from the nucleotide sequences in any one of **TABLE 6 and TABLE 5**.

[432] In some embodiments, systems comprise an effector protein described herein, a guide nucleic acid described herein, a reagent, support medium, or a combination thereof. In some embodiments, the effector protein comprises an effector protein, or a fusion protein thereof, described herein. In some embodiments, effector protein comprises an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of the amino acid sequences of **TABLE 1A** and **TABLE 1B**. In some embodiments, the amino acid sequence of the effector protein is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to any one of the amino acid sequences of **TABLE 1A** and **TABLE 1B**. In some embodiments, the guide nucleic acid comprises a repeat sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, or 100% identical to the nucleotide sequence set forth in **TABLE 6**.

[433] Disclosed herein, in some aspects, are systems for introduction of a donor nucleic acid into a safe harbor comprising one or more components. In some embodiments, the one or more components individually comprises one or more of the following: (i) an effector protein, or a nucleic acid encoding the effector protein; (ii) a guide nucleic acid, or a nucleic acid encoding the guide nucleic acid; and (iii) the donor nucleic acid encoding a transgene that comprises a functional human protein that is expressed

upon incorporation into the human safe harbor locus. In some embodiments, the effector protein comprises an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% sequence identity to any one of amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the guide nucleic acid comprises a spacer sequence that hybridizes to a target sequence in a target nucleic acid. In some embodiments, the target sequence comprises a nucleotide sequence within the safe harbor. In some embodiments, the safe harbor comprises at least 90% sequence identity to any one of the nucleotide sequences recited in **TABLE 10**.

[434] In some embodiments, the guide nucleic acid comprises at least one nucleotide sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to the nucleotide sequence of any one of **TABLE 4, TABLE 5, TABLE 6, TABLE 7, TABLE 8,** and **TABLE 9**. In some embodiments, the guide nucleic acid comprises at least one nucleotide sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to **SEQ ID NO: 1825, 1828, 1846** or **1847**. In some embodiments, the target nucleic acid comprises a nucleotide sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to a nucleotide sequence recited in **TABLE 10.** In some embodiments, the donor nucleic acid comprises a nucleotide sequence encoding amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to the amino acid sequence recited in **TABLE 11**.

[435] In some embodiments, systems comprise an effector protein described herein, a guide nucleic acid described herein, a reagent, support medium, or a combination thereof. In some embodiments, the effector protein comprises an effector protein, or a fusion protein thereof, described herein. In some embodiments, effector protein comprises an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to the amino acid sequence recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the amino acid sequence of the effector protein is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identical to the amino acid sequence recited in **TABLE 1A** and **TABLE 1B**. In some embodiments, the guide nucleic acid comprises a spacer sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, or 100% identical to any one of the nucleotide sequences recited in **TABLE 4** and **TABLE 5** and a repeat sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, or 100% identical to any one of the nucleotide sequences recited in **TABLE 6**. In some embodiments, the guide nucleic acid comprises a spacer sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, or 100% identical to any one of the nucleotide sequences recited in **TABLE 4** and **TABLE 5** and a handle sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, or 100% identical to any one of the nucleotide sequences recited in **TABLE 7**. In some embodiments, the nucleotide sequence of the guide nucleic acid is at least 70%, at least 75%, at

least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the gRNA sequences recited in **TABLE 9**.

[436] In some embodiments, systems described herein comprise separate compositions, solutions, containers, kits, vectors, or the like that individually comprise an effector protein, a nucleic acid encoding the effector protein, a guide nucleic acid, a nucleic acid encoding the guide nucleic acid, donor nucleic acid, or a combination thereof. Such systems can provide for separate delivery of the effector protein, the nucleic acid encoding the effector protein, the guide nucleic acid, the nucleic acid encoding the guide nucleic acid, or the donor nucleic acid described herein. In some embodiments, systems described herein comprise a composition, solution, container, kit, vector, or the like that comprises two or more of an effector protein, a nucleic acid encoding the effector protein, a guide nucleic acid, a nucleic acid encoding the guide nucleic acid, and donor nucleic acid described herein. Such systems can provide for delivery of two or more of the effector protein, the nucleic acid encoding the effector protein, the guide nucleic acid, the nucleic acid encoding the guide nucleic acid, and the donor nucleic acid.

Additional System Components

[437] In some embodiments, systems include a package, carrier, or container that is compartmentalized to receive one or more containers such as vials, tubes, and the like, each of the container(s) comprising one of the separate elements to be used in a method described herein. In some embodiments, any two system components are provided in different solutions or containers. Suitable containers include, for example, test wells, bottles, vials, and test tubes. In one embodiment, the containers are formed from a variety of materials such as glass, plastic, or polymers. The system or systems described herein contain packaging materials. Examples of packaging materials include, but are not limited to, pouches, blister packs, bottles, tubes, bags, containers, bottles, and any packaging material suitable for intended mode of use.

[438] A system may include labels listing contents and/or instructions for use, or package inserts with instructions for use. A set of instructions will also typically be included. In one embodiment, a label is on or associated with the container. In some embodiments, a label is on a container when letters, numbers or other characters forming the label are attached, molded, or etched into the container itself; a label is associated with a container when it is present within a receptacle or carrier that also holds the container, *e.g.*, as a package insert. In one embodiment, a label is used to indicate that the contents are to be used for a specific therapeutic application. The label also indicates directions for use of the contents, such as in the methods described herein. After packaging the formed product and wrapping or boxing to maintain a sterile barrier, the product may be terminally sterilized by heat sterilization, gas sterilization, gamma irradiation, or by electron beam sterilization. Alternatively, the product may be prepared and packaged by aseptic processing.

Amplification Reagents/Components

[439] In some embodiments, systems described herein comprise a reagent or component for amplifying a nucleic acid. Non-limiting examples of reagents for amplifying a nucleic acid include polymerases, primers, and nucleotides. In some embodiments, systems comprise reagents for nucleic acid amplification of a target nucleic acid in a sample. Nucleic acid amplification of the target nucleic acid may improve at least one of sensitivity, specificity, or accuracy of the assay in detecting the target nucleic acid. In some embodiments, nucleic acid amplification is isothermal nucleic acid amplification, providing for the use of the system or system in remote regions or low resource settings without specialized equipment for amplification. In some embodiments, amplification of the target nucleic acid increases the concentration of the target nucleic acid in the sample relative to the concentration of nucleic acids that do not correspond to the target nucleic acid.

[440] The reagents for nucleic acid amplification may comprise a recombinase, an oligonucleotide primer, a single-stranded DNA binding (SSB) protein, a polymerase, or a combination thereof that is suitable for an amplification reaction. Non-limiting examples of amplification reactions are transcription mediated amplification (TMA), helicase dependent amplification (HDA), or circular helicase dependent amplification (cHDA), strand displacement amplification (SDA), recombinase polymerase amplification (RPA), loop mediated amplification (LAMP), exponential amplification reaction (EXPAR), rolling circle amplification (RCA), ligase chain reaction (LCR), simple method amplifying RNA targets (SMART), single primer isothermal amplification (SPIA), multiple displacement amplification (MDA), nucleic acid sequence based amplification (NASBA), hinge-initiated primer-dependent amplification of nucleic acids (HIP), nicking enzyme amplification reaction (NEAR), and improved multiple displacement amplification (IMDA).

[441] In some embodiments, systems comprise a PCR tube, a PCR well or a PCR plate. The wells of the PCR plate may be pre-aliquoted with the reagent for amplifying a nucleic acid, as well as a guide nucleic acid, an effector protein, a multimeric complex, or any combination thereof. The wells of the PCR plate may be pre-aliquoted with a guide nucleic acid targeting a target sequence, an effector protein capable of being activated when complexed with the guide nucleic acid and the target sequence.

[442] In some embodiments, systems comprise a support medium; a guide nucleic acid targeting a target sequence; and an effector protein capable of being activated when complexed with the guide nucleic acid and the target sequence. In some embodiments, nucleic acid amplification is performed in a nucleic acid amplification region on the support medium. Alternatively, or in combination, the nucleic acid amplification is performed in a reagent chamber, and the resulting sample is applied to the support medium.

[443] In some embodiments, a system for modifying a target nucleic acid comprises a PCR plate; a guide nucleic acid targeting a target sequence; and an effector protein capable of being activated when complexed with the guide nucleic acid and the target sequence. The wells of the PCR plate may be pre-aliquoted with the guide nucleic acid targeting a target sequence, and an effector protein capable of

being activated when complexed with the guide nucleic acid and the target sequence. A user may thus add the biological sample of interest to a well of the pre-aliquoted PCR plate.

[444] Often, the nucleic acid amplification is performed for no greater than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, or 60 minutes, or any value 1 to 60 minutes. Sometimes, the nucleic acid amplification is performed for 1 to 60, 5 to 55, 10 to 50, 15 to 45, 20 to 40, or 25 to 35 minutes. Sometimes, the nucleic acid amplification reaction is performed at a temperature of around 20-45°C. In some embodiments, the nucleic acid amplification reaction is performed at a temperature no greater than 20°C, 25°C, 30°C, 35°C, 37°C, 40°C, 45°C, or any value 20 °C to 45 °C. In some embodiments, the nucleic acid amplification reaction is performed at a temperature of at least 20°C, 25°C, 30°C, 35°C, 37°C, 40°C, or 45°C, or any value 20 °C to 45 °C. In some embodiments, the nucleic acid amplification reaction is performed at a temperature of 20°C to 45°C, 25°C to 40°C, 30°C to 40°C, or 35°C to 40°C.

Certain System Conditions

[445] Certain conditions that may enhance the activity of an effector protein include a certain salt presence or salt concentration of the solution in which the activity occurs. For example, cis-cleavage activity of an effector protein may be inhibited or halted by a high salt concentration. The salt may be a sodium salt, a potassium salt, or a magnesium salt. In some embodiments, the salt is NaCl. In some embodiments, the salt is KNO_3. In some embodiments, the salt concentration is less than 150 mM, less than 125 mM, less than 100 mM, less than 75 mM, less than 50 mM, or less than 25 mM.

[446] Certain conditions that may enhance the activity of an effector protein include the pH of a solution in which the activity. For example, increasing pH may enhance trans cleavage activity. For example, the rate of trans cleavage activity may increase with increase in pH up to pH 9. In some embodiments, the pH is about 7, about 7.1, about 7.2, about 7.3, about 7.4, about 7.5, about 7.6, about 7.7, about 7.8, about 7.9, about 8, about 8.1, about 8.2, about 8.3, about 8.4, about 8.5, about 8.6, about 8.7, about 8.8, about 8.9, or about 9. In some embodiments, the pH is 7 to 7.5, 7.5 to 8, 8 to 8.5, 8.5 to 9, or 7 to 8.5. In some embodiments, the pH is less than 7. In some embodiments, the pH is greater than 7.

[447] Certain conditions that may enhance the activity of an effector protein includes the temperature at which the activity is performed. In some embodiments, the temperature is about 25°C to about 50°C. In some embodiments, the temperature is about 20°C to about 40°C, about 30°C to about 50°C, or about 40°C to about 60°C. In some embodiments, the temperature is about 25°C, about 30°C, about 35°C, about 40°C, about 45°C, or about 50°C.

X. Methods and Formulations for Introducing System Components and Compositions into a Target Cell

[448] A guide nucleic acid (or a nucleic acid comprising a nucleotide sequence encoding same) and/or an effector protein described herein may be introduced into a host cell by any of a variety of well-known methods. As a non-limiting example, a guide nucleic acid and/or effector protein may be combined with a lipid. As another non-limiting example, a guide nucleic acid and/or effector protein may be combined with a particle or formulated into a particle.

[449] Described herein are methods of introducing various components described herein to a host. A host may be any suitable host, such as a host cell. When described herein, a host cell may be an *in vivo* or *in vitro* eukaryotic cell, a prokaryotic cell (*e.g.*, bacterial or archaeal cell), or a cell from a multicellular organism (*e.g.*, a cell line) cultured as a unicellular entity, which eukaryotic or prokaryotic cells may be, or have been, used as recipients for methods of introduction described herein, and include the progeny of the original cell which has been transformed by the methods of introduction described herein. It is understood that the progeny of a single cell may not necessarily be completely identical in morphology or in genomic or total DNA complement as the original parent, due to natural, accidental, or deliberate mutation. A host cell may be a recombinant host cell or a genetically modified host cell, if a heterologous nucleic acid, *e.g.*, an expression vector, has been introduced into the cell.

[450] Methods of introducing a nucleic acid and/or protein into a host cell are known in the art, and any convenient method may be used to introduce a subject nucleic acid (*e.g.*, an expression construct/vector) into a target cell (*e.g.*, a human cell, and the like). Suitable methods include, *e.g.*, viral infection, transfection, conjugation, protoplast fusion, lipofection, electroporation, calcium phosphate precipitation, polyethyleneimine (PEI)-mediated transfection, DEAE-dextran mediated transfection, liposome-mediated transfection, particle gun technology, calcium phosphate precipitation, direct micro injection, nanoparticle-mediated nucleic acid delivery (see, *e.g.*, Panyam *et al.* Adv Drug Deliv Rev. 2012 Sep 13. pii: S0169-409X(12)00283-9. doi: 10.1016/j.addr.2012.09.023), and the like. In some embodiments, the nucleic acid and/or protein are introduced into a disease cell comprised in a pharmaceutical composition comprising the guide nucleic acid and/or effector protein and a pharmaceutically acceptable excipient.

[451] In some embodiments, molecules of interest, such as nucleic acids of interest, are introduced to a host. In some embodiments, polypeptides, such as an effector protein are introduced to a host. In some embodiments, vectors, such as lipid particles and/or viral vectors may be introduced to a host. Introduction may be for contact with a host or for assimilation into the host, for example, introduction into a host cell.

[452] In some embodiments, described herein are methods of introducing one or more nucleic acids, such as a nucleic acid encoding an effector protein, a nucleic acid that, when transcribed, produces an engineered guide nucleic acid, and/or a donor nucleic acid, or combinations thereof, into a host cell. Any suitable method may be used to introduce a nucleic acid into a cell. Suitable methods include, for example, viral infection, transfection, lipofection, electroporation, calcium phosphate precipitation, polyethyleneimine (PEI)-mediated transfection, DEAE-dextran mediated transfection, liposome-

mediated transfection, particle gun technology, calcium phosphate precipitation, direct microinjection, nanoparticle-mediated nucleic acid delivery, and the like. Further methods are described throughout.

[453] Introducing one or more nucleic acids into a host cell may occur in any culture media and under any culture conditions that promote the survival of the cells. Introducing one or more nucleic acids into a host cell may be carried out *in vivo* or *ex vivo*. Introducing one or more nucleic acids into a host cell may be carried out *in vitro*.

[454] In some embodiments, an effector protein may be provided as RNA. The RNA may be provided by direct chemical synthesis or may be transcribed *in vitro* from a DNA (*e.g.*, encoding the effector protein). Once synthesized, the RNA may be introduced into a cell by way of any suitable technique for introducing nucleic acids into cells (*e.g.*, microinjection, electroporation, transfection, *etc.*). In some embodiments, introduction of one or more nucleic acid may be through the use of a vector and/or a vector system, accordingly, in some embodiments, compositions and system described herein comprise a vector and/or a vector system.

[455] Vectors may be introduced directly to a host. In some embodiments, host cells may be contacted with one or more vectors as described herein, and in some embodiments, said vectors are taken up by the cells. Methods for contacting cells with vectors include but are not limited to electroporation, calcium chloride transfection, microinjection, lipofection, micro-injection, contact with the cell or particle that comprises a molecule of interest, or a package of cells or particles that comprise molecules of interest.

[456] Components described herein may also be introduced directly to a host. For example, an engineered guide nucleic acid may be introduced to a host, specifically introduced into a host cell. Methods of introducing nucleic acids, such as RNA into cells include, but are not limited to direct injection, transfection, or any other method used for the introduction of nucleic acids.

[457] Polypeptides (*e.g.*, effector proteins) described herein may also be introduced directly to a host. In some embodiments, polypeptides described herein may be modified to promote introduction to a host. For example, polypeptides described herein may be modified to increase the solubility of the polypeptide. Such a polypeptide may optionally be fused to a polypeptide domain that increases solubility. The domain may be linked to the polypeptide through a defined protease cleavage site, such as TEV sequence which is cleaved by TEV protease. The linker may also include one or more flexible sequences, *e.g.* from 1 to 10 glycine residues. In some embodiments, the cleavage of the polypeptide is performed in a buffer that maintains solubility of the product, *e.g.* in the presence of from 0.5 to 2 M urea, in the presence of polypeptides and/or polynucleotides that increase solubility, and the like. Domains of interest include endosomolytic domains, *e.g.* influenza HA domain; and other polypeptides that aid in production, *e.g.* IF2 domain, GST domain, GRPE domain, and the like. In another example, the polypeptide may be modified to improve stability. For example, the polypeptides may be PEGylated, where the polyethyleneoxy group provides for enhanced lifetime in the blood stream. Polypeptides may also be modified to promote uptake by a host, such as a host cell. For example, a

polypeptide described herein may be fused to a polypeptide permeant domain to promote uptake by a host cell. Any suitable permeant domains may be used in the non-integrating polypeptides of the present disclosure, including peptides, peptidomimetics, and non-peptide carriers. Examples include penetratin, a permeant peptide may be derived from the third alpha helix of Drosophila melanogaster transcription factor Antennapaedia; the HIV-1 tat basic region amino acid sequence, *e.g.*, amino acids 49-57 of a naturally-occurring tat protein; and poly-arginine motifs, for example, the region of amino acids 34-56 of HIV-1 rev protein, nonaarginine, octa-arginine, and the like. The site at which the fusion is made may be selected in order to optimize the biological activity, secretion or binding characteristics of the polypeptide. The optimal site may be determined by suitable methods.

[458] Described herein are formulations of introducing compositions or components of a system described herein to a host. In some embodiments, such formulations, systems and compositions described herein comprise an effector protein and a carrier (*e.g.*, excipient, diluent, vehicle, or filling agent). In some aspects of the present disclosure, the effector protein is provided in a pharmaceutical composition comprising the effector protein and any pharmaceutically acceptable excipient, carrier, or diluent.

XI. Methods of Nucleic Acid Editing

[459] Provided herein are methods of editing target nucleic acids. In general, editing refers to modifying the nucleotide sequence of a target nucleic acid. However, compositions and systems disclosed herein may also be capable of making epigenetic modifications of target nucleic acids. Effector proteins, multimeric complexes thereof and systems described herein may be used for editing or modifying a target nucleic acid. Editing a target nucleic acid may comprise one or more of cleaving the target nucleic acid, deleting one or more nucleotides of the target nucleic acid, inserting one or more nucleotides into the target nucleic acid, mutating one or more nucleotides of the target nucleic acid, or modifying (*e.g.*, methylating, demethylating, deaminating, or oxidizing) of one or more nucleotides of the target nucleic acid.

[460] Methods of editing may comprise contacting a target nucleic acid with an effector protein described herein and a guide nucleic acid, wherein the effector protein comprises an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the amino acid sequences set forth in **TABLE 1A** and **TABLE 1B**. In some embodiments, the guide nucleic acid comprises a repeat sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, or 100% identical to the nucleotide sequence set forth in **TABLE 6**. In some embodiments, the guide nucleic acid comprises a spacer sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, or 100% identical to any one of the nucleotide sequences recited in **TABLE 4** and **TABLE 5** and a repeat sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, or 100% identical to the nucleotide sequence recited

in **TABLE 6**. In some embodiments, the guide nucleic acid comprises a spacer sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, or 100% identical to any one of the nucleotide sequences recited in **TABLE 4** and **TABLE 5** and a handle sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, or 100% identical to any one of the nucleotide sequences recited in **TABLE 7**. In some embodiments, the nucleotide sequence of the guide nucleic acid is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the nucleotide sequences recited in **TABLE 4, TABLE 5, TABLE 6, TABLE 7, TABLE 8,** and **TABLE 9**. In some embodiments, the nucleotide sequence of the guide nucleic acid is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of **SEQ ID NO: 1825, 1828, 1846** or **1847**.

[461] In some embodiments, the compositions, methods or systems comprise a nucleic acid expression vector, or use thereof, to introduce an effector protein, guide nucleic acid, donor template or any combination thereof to a cell. In some embodiments, the nucleic acid expression vector is a viral vector. Viral vectors include, but are not limited to, retroviruses, adenoviruses, adeno-associated viruses, and herpes simplex viruses. In some embodiments, the viral vector is a replication-defective viral vector, comprising an insertion of a therapeutic gene inserted in genes essential to the lytic cycle, preventing the virus from replicating and exerting cytotoxic effects. In some embodiments, the viral vector is an adeno associated viral (AAV) vector. In some embodiments, the nucleic acid expression vector is a non-viral vector. In some embodiments, compositions and methods comprise a lipid, polymer, nanoparticle, or a combination thereof, or use thereof, to introduce a Cas protein, guide nucleic acid, donor template or any combination thereof to a cell. Non-limiting examples of lipids and polymers are cationic polymers, cationic lipids, or bio-responsive polymers. In some embodiments, the bio-responsive polymer exploits chemical-physical properties of the endosomal environment (*e.g.*, pH) to preferentially release the genetic material in the intracellular space.

[462] Methods of editing may comprise contacting a target nucleic acid with one or more components, compositions or systems described herein. In some embodiments, a method of editing comprises contacting a target nucleic acid with at least one of: a) one or more effector proteins, or one or more nucleic acids encoding one or more effector proteins; or b) one or more guide nucleic acids, or one or more nucleic acids encoding one or more guide nucleic acids. In some embodiments, a method of editing comprises contacting a target nucleic acid with a system described herein wherein the system comprises components comprising at least one of: a) one or more effector proteins, or one or more nucleic acids encoding one or more effector proteins; or b) one or more guide nucleic acids, or one or more nucleic acids encoding one or more guide nucleic acids. In some embodiments, a method of editing comprises contacting a target nucleic acid with a composition described herein comprising at least one of: a) one or more effector proteins, or one or more nucleic acids encoding one or more effector

proteins; or b) one or more guide nucleic acids, or one or more nucleic acids encoding one or more guide nucleic acids; in a composition.

[463] Editing may introduce a mutation (*e.g.*, point mutations, insertions, deletions) in a target nucleic acid relative to a corresponding wildtype nucleotide sequence. Editing may remove or insert a nucleic acid sequence to produce a corresponding wildtype protein. Editing may remove/insert tissue-specific nucleic acid sequence in a target nucleic acid. Editing may be used to generate gene knock-in, gene editing, or a combination thereof. Methods of the disclosure may be targeted to any locus in a genome of a cell.

[464] Editing may comprise single stranded cleavage, double stranded cleavage, donor nucleic acid insertion, epigenetic modification (*e.g.*, methylation, demethylation, acetylation, or deacetylation), or a combination thereof. In some embodiments, cleavage (single-stranded or double-stranded) is site-specific, meaning cleavage occurs at a specific site in the target nucleic acid, often within the region of the target nucleic acid that hybridizes with the guide nucleic acid spacer region. In some embodiments, the effector proteins introduce a single-stranded break in a target nucleic acid to produce a cleaved nucleic acid. In some embodiments, the effector protein is capable of introducing a break in a single stranded RNA (ssRNA). The effector protein may be coupled to a guide nucleic acid that targets a particular region of interest in the ssRNA. In some embodiments, the target nucleic acid, and the resulting cleaved nucleic acid is contacted with a nucleic acid for homologous recombination (*e.g.*, homology directed repair (HDR)) or non-homologous end joining (NHEJ). In some embodiments, a double-stranded break in the target nucleic acid may be repaired (*e.g.*, by NHEJ or HDR) without insertion of a donor template, such that the repair results in an indel in the target nucleic acid at or near the site of the double-stranded break.

[465] In some embodiments, an indel, sometimes referred to as an insertion-deletion or indel mutation, is a type of genetic mutation that results from the insertion and/or deletion of nucleotides in a target nucleic acid. An indel can vary in length (*e.g.*, 1 to 1,000 nucleotides in length) and be detected using methods well known in the art, including sequencing. If the number of nucleotides in the insertion/deletion is not divisible by three, and it occurs in a protein coding region, it is also a frameshift mutation.

[466] In some embodiments, an indel percentage is based on a percentage of sequencing reads that show at least one nucleotide has been edited from the insertion and/or deletion of nucleotides regardless of the size of insertion or deletion, or number of nucleotides edited. For example, if there is at least one nucleotide deletion detected in a given target nucleic acid, it counts towards the percent indel value. As another example, if one copy of the target nucleic acid has one nucleotide deleted, and another copy of the target nucleic acid has 10 nucleotides deleted, they are counted the same. This number reflects the percentage of target nucleic acids that are edited by a given effector protein.

[467] In some embodiments, methods of editing described herein cleave a target nucleic acid at one or more locations to generate a cleaved target nucleic acid. In some embodiments, the cleaved target

nucleic acid undergoes recombination (*e.g.*, NHEJ or HDR). In some embodiments, cleavage in the target nucleic acid may be repaired (*e.g.*, by NHEJ or HDR) with insertion of a donor nucleic acid, such that the repair results in an indel in the target nucleic acid at or near the site of the cleavage site.

[468] In some embodiments, wherein the compositions, systems, and methods of the present disclosure comprise an additional guide nucleic acid or a use thereof, the dual-guided compositions, systems, and methods described herein can modify the target nucleic acid in two locations. In some embodiments, dual-guided editing can comprise cleavage of the target nucleic acid in the two locations targeted by the guide RNAs. In certain embodiments, upon removal of the sequence between the guide nucleic acids, a new nucleotide sequence can be inserted.

[469] Accordingly, in some embodiments, compositions, systems, and methods described herein can edit 1 to 1,000 nucleotides or any integer in between, in a target nucleic acid. In certain embodiments, 1 to 1,000, 2 to 900, 3 to 800, 4 to 700, 5 to 600, 6 to 500, 7 to 400, 8 to 300, 9 to 200, or 10 to 100 nucleotides, or any integer in between, can be edited by the compositions, systems, and methods described herein. In some embodiments, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more nucleotides can be edited by the compositions, systems, and methods described herein. In some embodiments, 10, 20, 30, 40, 50, 60, 70, 80 90, 100 or more nucleotides, or any integer in between, can be edited by the compositions, systems, and methods described herein. In some embodiments, 100, 200, 300, 400, 500, 600, 700, 800, 900 or more nucleotides, or any integer in between, can be edited by the compositions, systems, and methods described herein.

[470] Methods may comprise use of two or more effector proteins. An illustrative method for introducing a break in a target nucleic acid comprises contacting the target nucleic acid with: (a) a first engineered guide nucleic acid comprising a region that binds to a first effector protein, wherein the effector protein comprises at least 75% sequence identity to the amino acid sequence of **TABLE 1A** and **TABLE 1B**; and (b) a second engineered guide nucleic acid comprising a region that binds to a second effector protein, wherein the effector protein comprises at least 75% sequence identity to the amino acid sequence of **TABLE 1A** and **TABLE 1B**, wherein the first engineered guide nucleic acid comprises an additional region that binds to the target nucleic acid and wherein the second engineered guide nucleic acid comprises an additional region that binds to the target nucleic acid. In some embodiments, the guide nucleic acid comprises a crRNA sequence comprising a repeat sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, or 100% identical to the nucleotide sequence of **TABLE 6**.

[471] In some embodiments, editing a target nucleic acid comprises genome editing. Genome editing may comprise modifying a genome, chromosome, plasmid, or other genetic material of a cell or organism. In some embodiments, the genome, chromosome, plasmid, or other genetic material of the cell or organism is modified in vivo. In some embodiments, the genome, chromosome, plasmid, or other genetic material of the cell or organism is modified in a cell. In some embodiments, the genome, chromosome, plasmid, or other genetic material of the cell or organism is modified in vitro. For

example, a plasmid may be modified in vitro using a composition described herein and introduced into a cell or organism. In some embodiments, modifying a target nucleic acid may comprise deleting a sequence from a target nucleic acid. In some embodiments, modifying a target nucleic acid may comprise replacing a sequence in a target nucleic acid with a second sequence. In some embodiments, modifying a target nucleic acid may comprise introducing a sequence into a target nucleic acid. For example, a beneficial sequence or a sequence that may reduce or eliminate a disease may be inserted into the target nucleic acid. In some embodiments, the beneficial sequence is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, or 100% identical to an amino acid sequence recited in **TABLE 11.**

[472] In some embodiments, methods comprise editing a target nucleic acid with two or more effector proteins. Editing a target nucleic acid may comprise introducing a two or more single-stranded breaks in a target nucleic acid. In some embodiments, a break may be introduced by contacting a target nucleic acid with an effector protein and a guide nucleic acid. The guide nucleic acid may bind to the effector protein and hybridize to a region of the target nucleic acid, thereby recruiting the effector protein to the region of the target nucleic acid. Binding of the effector protein to the guide nucleic acid and the region of the target nucleic acid may activate the effector protein, and the effector protein may introduce a break (*e.g.,* a single stranded break) in the region of the target nucleic acid. In some embodiments, modifying a target nucleic acid may comprise introducing a first break in a first region of the target nucleic acid and a second break in a second region of the target nucleic acid. For example, modifying a target nucleic acid may comprise contacting a target nucleic acid with a first guide nucleic acid that binds to a first effector protein and hybridizes to a first region of the target nucleic acid and a second guide nucleic acid that binds to a second programmable nickase and hybridizes to a second region of the target nucleic acid. The first effector protein may introduce a first break in a first strand at the first region of the target nucleic acid, and the second effector protein may introduce a second break in a second strand at the second region of the target nucleic acid. In some embodiments, a segment of the target nucleic acid between the first break and the second break may be removed, thereby modifying the target nucleic acid. In some embodiments, a segment of the target nucleic acid between the first break and the second break may be replaced (*e.g.,* with donor nucleic acid), thereby modifying the target nucleic acid. In some embodiments, the effector protein comprises an amino acid sequence that is at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the amino acid sequences of **TABLE 1A** and **TABLE 1B**.

[473] In some embodiments, methods comprise inserting a donor nucleic acid into a cleaved target nucleic acid. The donor nucleic acid may be inserted at a specified (*e.g.*, effector protein targeted) point within the target nucleic acid. In some embodiments, methods comprise contacting a target nucleic acid with an effector protein comprising an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical

to the amino acid sequence of **TABLE 1A** and **TABLE 1B**, thereby introducing a single-stranded break in the target nucleic acid; contacting the target nucleic acid with a second effector protein, optionally comprising an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, or at least 99% identical to the amino acid sequence of **TABLE 1A** and **TABLE 1B**, to generate a second cleavage site in the target nucleic acid, ligating the regions flanking the first and second cleavage site, optionally through NHEJ or single-strand annealing, thereby resulting in the excision of a portion of the target nucleic acid between the first and second cleavage sites from the target nucleic acid; and contacting the target nucleic acid with a donor nucleic acid for homologous recombination, optionally via HDR or NHEJ, thereby introducing a new sequence into the target nucleic acid (*e.g.*, at a cleavage site or in between two cleavage sites).

[474] In some embodiments, methods comprise editing a target nucleic acid with two or more effector proteins. Editing a target nucleic acid may comprise introducing a two or more single-stranded breaks in a target nucleic acid. In some embodiments, a break may be introduced by contacting a target nucleic acid with an effector protein and a guide nucleic acid. The guide nucleic acid may bind to the effector protein and hybridize to a region of the target nucleic acid, thereby recruiting the effector protein to the region of the target nucleic acid. Binding of the effector protein to the guide nucleic acid and the region of the target nucleic acid may activate the effector protein, and the effector protein may introduce a break (*e.g.*, a single stranded break) in the region of the target nucleic acid. In some embodiments, modifying a target nucleic acid may comprise introducing a first break in a first region of the target nucleic acid and a second break in a second region of the target nucleic acid. For example, modifying a target nucleic acid may comprise contacting a target nucleic acid with a first guide nucleic acid that binds to a first effector protein and hybridizes to a first region of the target nucleic acid and a second guide nucleic acid that binds to a second programmable nickase and hybridizes to a second region of the target nucleic acid. The first effector protein may introduce a first break in a first strand at the first region of the target nucleic acid, and the second effector protein may introduce a second break in a second strand at the second region of the target nucleic acid. In some embodiments, a segment of the target nucleic acid between the first break and the second break may be removed, thereby modifying the target nucleic acid. In some embodiments, a segment of the target nucleic acid between the first break and the second break may be replaced (*e.g.*, with donor nucleic acid), thereby modifying the target nucleic acid. In some embodiments, the effector protein comprises an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to the amino acid sequence of **TABLE 1A** and **TABLE 1B**. In some embodiments, the guide nucleic acid comprises a crRNA sequence comprising a repeat sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, or 100% identical to the nucleotide sequence of **TABLE 6**.

[475] In some embodiments, editing is achieved by fusing an effector protein to a heterologous sequence. The heterologous sequence may be a suitable fusion partner, *e.g.*, a protein that provides

recombinase activity by acting on the target nucleic acid. In some embodiments, the fusion protein comprises an effector protein fused to a heterologous sequence by a linker. The heterologous sequence or fusion partner may be a base editing domain. The base editing domain may be an ADAR1/2 or any functional variant thereof. The heterologous sequence or fusion partner may be fused to the C-terminus, N-terminus, or an internal portion (*e.g.*, a portion other than the N- or C-terminus) of the effector protein. The heterologous sequence or fusion partner may be fused to the effector protein by a linker. A linker may be a peptide linker or a non-peptide linker. In some embodiments, the linker is an XTEN linker. In some embodiments, the linker comprises one or more repeats a tri-peptide GGS. In some embodiments, the linker is from 1 to 100 amino acids in length. In some embodiments, the linker is more 100 amino acids in length. In some embodiments, the linker is from 10 to 27 amino acids in length. A non-peptide linker may be a polyethylene glycol (PEG), polypropylene glycol (PPG), co-poly(ethylene/propylene) glycol, polyoxyethylene (POE), polyurethane, polyphosphazene, polysaccharides, dextran, polyvinyl alcohol, polyvinylpyrrolidones, polyvinyl ethyl ether, polyacrylamide, polyacrylate, polycyanoacrylates, lipid polymers, chitins, hyaluronic acid, heparin, or an alkyl linker.

[476] In certain embodiments, editing or modification of a target nucleic acid can be locus specific, wherein compositions, systems, and methods described herein can edit or modify a target nucleic acid at one or more specific loci to effect one or more specific mutations comprising sequence deletion, sequence knock-in, or any combination thereof. For example, editing or modification of a specific locus can effect sequence knock-in. In certain embodiments, sequence knock-in is a modification where one or more sequences is inserted into a target nucleic acid relative to a target nucleic acid without the sequence knock-in. In certain embodiments, editing or modification of a specific locus can effect sequence knock-in and sequence deletion. In certain embodiments, editing or modification of a target nucleic acid can be locus specific, modification specific, or both. In certain embodiments, editing or modification of a target nucleic acid can be locus specific, modification specific, or both, wherein compositions, systems, and methods described herein comprise an effector protein described herein, and a guide nucleic acid described herein. In certain embodiments, edition or modification of a target nucleic acid is specific to intron 1 of mammalian albumin gene.

[477] Methods of editing a target nucleic acid or modulating the expression of a target nucleic acid may be performed *in vivo*. Methods of editing a target nucleic acid or modulating the expression of a target nucleic acid may be performed *in vitro*. For example, a plasmid may be modified *in vitro* using a composition described herein and introduced into a cell or organism. Methods of editing a target nucleic acid or modulating the expression of a target nucleic acid may be performed *ex vivo*. For example, methods may comprise obtaining a cell from a subject, modifying a target nucleic acid in the cell with methods described herein, and returning the cell to the subject.

Transfection Donor Nucleic Acids

[478] In reference to a viral vector, the term transfection donor nucleic acid refers to a sequence of nucleotides that will be or has been introduced into a cell following transfection of the viral vector. The transfection donor nucleic acid may be introduced into the cell by any mechanism of the transfecting viral vector, including, but not limited to, integration into the genome of the cell or introduction of an episomal plasmid or viral genome. As another example, when used in reference to the activity of an effector protein, the term donor nucleic acid refers to a sequence of nucleotides that will be or has been inserted at the site of cleavage by the effector protein (cleaving (hydrolysis of a phosphodiester bond) of a nucleic acid resulting in a nick or double strand break –nuclease activity). As yet another example, when used in reference to homologous recombination, the term donor nucleic acid refers to a sequence of DNA that serves as a template in the process of homologous recombination, which may carry the modification that is to be or has been introduced into the target nucleic acid. By using this donor nucleic acid as a template, the genetic information, including the modification, is copied into the target nucleic acid by way of homologous recombination.

[479] Donor nucleic acids of any suitable size may be integrated into a target nucleic acid or genome. In some embodiments, the donor polynucleotide integrated into a genome is less than 3, about 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10, 10.5, 11, 11.5, 12, 12.5, 13, 13.5, 14, 14.5, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500 kilobases in length. In some embodiments, donor nucleic acids are more than 500 kilobases (kb) in length.

[480] The donor nucleic acid may comprise a sequence that is derived from an animal. The animal may be human. The animal may be a non-human animal, such as, by way of non-limiting example, a mouse, rat, hamster, rabbit, pig, bovine, deer, sheep, goat, chicken, cat, dog, ferret, a bird, non-human primate (*e.g.*, marmoset, rhesus monkey). The non-human animal may be a domesticated mammal or an agricultural mammal.

[481] In some embodiments, a viral vector comprising a donor nucleic acid introduces the donor nucleic acid into a cell following transfection. In some embodiments, the donor nucleic acid is introduced into the cell by any mechanism of the transfecting viral vector, including, but not limited to, integration into the genome of the cell or introduction of an episomal plasmid or viral genome.

[482] In some embodiments, an effector protein as described herein facilitates insertion of a donor nucleic acid at a site of cleavage or between two cleavage sites by cleaving (hydrolysis of a phosphodiester bond) of a nucleic acid resulting in a nick or double strand break – nuclease activity.

[483] In some embodiments, a donor nucleic acid serves as a template in the process of homologous recombination, which may carry a modification that is to be or has been introduced into a target nucleic acid. By using the donor nucleic acid as a template, the genetic information, including the modification, is copied into the target nucleic acid by way of homologous recombination.

Genetically Modified Cells and Organisms

[484] Methods of editing described herein may be employed to generate a genetically modified cell. The cell may be a eukaryotic cell (*e.g.*, a mammalian cell) or a prokaryotic cell (*e.g.*, an archaeal cell). The cell may be derived from a multicellular organism and cultured as a unicellular entity. The cell may comprise a heritable genetic modification, such that progeny cells derived therefrom comprise the heritable genetic mutation. The cell may be progeny of a genetically modified cell comprising a genetic modification of the genetically modified parent cell. A genetically modified cell may comprise a deletion, insertion, mutation, or non-native sequence relative to a wild-type version of the cell or the organism from which the cell was derived.

[485] In some embodiments, upon modification of a target nucleic acid by compositions, systems, and methods described herein, the target nucleic acid can comprise an intron deletion, intron knock-in, or a combination thereof. In some embodiments, a cell modified by any one of the compositions described herein, any one of the systems described herein, or any one of the methods described herein.

[486] Methods may comprise contacting a cell with a nucleic acid (*e.g.*, a plasmid or mRNA) comprising a nucleotide sequence encoding an effector protein, wherein the effector protein comprises an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, or at least 99% identical to any one of the amino acid sequences of **TABLE 1A** and **TABLE 1B**.

[487] Methods may comprise contacting cells with a nucleic acid (*e.g.*, a plasmid or mRNA) comprising a nucleotide sequence encoding an effector protein, a guide nucleic acid (*e.g.*, a crRNA or sgRNA), a donor nucleic acid or any combination thereof. Methods may comprise contacting cells with a nucleic acid (*e.g.*, a plasmid or mRNA) comprising a nucleotide sequence encoding a guide nucleic acid, a tracrRNA, a crRNA, or any combination thereof. Contacting may comprise electroporation, acoustic poration, optoporation, viral vector-based delivery, iTOP, nanoparticle delivery (*e.g.*, lipid or gold nanoparticle delivery), cell-penetrating peptide (CPP) delivery, DNA nanostructure delivery, or any combination thereof.

[488] Methods may comprise contacting cells with a nucleic acid (*e.g.*, a plasmid or mRNA) comprising a nucleotide sequence encoding an effector protein, a guide nucleic acid (*e.g.*, a crRNA or sgRNA), a donor nucleic acid or any combination thereof. In some embodiments, methods may comprise contacting cells with a nucleic acid (*e.g.*, a plasmid or mRNA) comprising a nucleotide sequence encoding a guide nucleic acid. In some embodiments, the nucleotide sequence of the guide nucleic acid is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to of any one of the gRNA sequences recited in **TABLE 4, TABLE 5, TABLE 6, TABLE 7, TABLE 8,** and **TABLE 9**. In some embodiments, the guide nucleic acid comprises a crRNA sequence comprising a spacer sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the spacer sequences recited in **TABLE 4** and **TABLE 5,** and a repeat sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at

least 90%, at least 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the repeat sequences recited in **TABLE 6**. In some embodiments, the guide nucleic acid comprises a crRNA sequence comprising a spacer sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the spacer sequences recited in **TABLE 4** and **TABLE 5,** and a handle sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the handle sequences recited in **TABLE 7**.

[489] Methods may comprise contacting cells with a nucleic acid (*e.g.,* a plasmid or mRNA) comprising a nucleotide sequence encoding an effector protein, a guide nucleic acid (*e.g.,* a crRNA or sgRNA), a donor nucleic acid or any combination thereof. Methods may comprise contacting cells with a nucleic acid (*e.g.,* a plasmid or mRNA) comprising a nucleotide sequence encoding a guide nucleic acid, a tracrRNA, a sgRNA, or any combination thereof. Contacting may comprise electroporation, acoustic poration, optoporation, viral vector-based delivery, iTOP, nanoparticle delivery (*e.g.,* lipid or gold nanoparticle delivery), cell-penetrating peptide (CPP) delivery, DNA nanostructure delivery, or any combination thereof.

[490] Methods may comprise contacting a cell with an effector protein or a multimeric complex thereof, wherein the effector protein comprises an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, or at least 99% identical to any one of the amino acid sequences of **TABLE 1A** and **TABLE 1B**.

[491] Methods of the disclosure may be performed in a subject. Compositions of the disclosure may be administered to a subject. A subject may be a human. A subject may be a mammal (*e.g.*, rat, mouse, cow, dog, pig, sheep, horse). A subject may be a vertebrate or an invertebrate. A subject may be a laboratory animal. A subject may be a patient. A subject may be at risk of developing, suffering from, or displaying symptoms of disease. In some embodiments, a mutation comprises a point mutation or single nucleotide polymorphism (SNP), a chromosomal mutation, a copy number mutation, or any combination thereof. A point mutation optionally comprises a substitution, insertion, or deletion. In some embodiments, a mutation comprises a chromosomal mutation. A chromosomal mutation can comprise an inversion, a deletion, a duplication, or a translocation. In some embodiments, a mutation comprises a copy number variation. A copy number variation can comprise a gene amplification or an expanding trinucleotide repeat.

[492] Methods of the disclosure may be performed in a cell. A cell may be in vitro. A cell may be in vivo. A cell may be ex vivo. A cell may be an isolated cell. A cell may be a cell inside of an organism. A cell may be an organism. A cell may be a cell in a cell culture. A cell may be one of a collection of cells. A cell may be a mammalian cell or derived from a mammalian cell. A cell may be a rodent cell or derived from a rodent cell. A cell may be a human cell or derived from a human cell. A cell may be a eukaryotic cell or derived from a eukaryotic cell. A cell may be a pluripotent stem cell. A cell may

be an animal cell or derived from an animal cell. A cell may be an invertebrate cell or derived from an invertebrate cell. A cell may be a vertebrate cell or derived from a vertebrate cell.

[493] A cell may be from a specific organ or tissue. A cell may be hepatocyte. The tissue may be the subject's blood, bone marrow, or cord blood. The tissue may be heterologous donor blood, cord blood, or bone marrow. The tissue may be allogenic blood, cord blood, or bone marrow. In some embodiments, the cell is a: a stem cell, progenitor cell, a pluripotent stem cell or a cell derived from a pluripotent stem cell.

XII. Methods of Treating Diseases

[494] Described herein are methods for treating a disease in a subject by modifying a target nucleic acid associated with a gene or expression of a gene related to the disease. In some embodiments, the disease or disorder comprises Pompe disease, glycogen storage disorder, cystic fibrosis, muscular dystrophy, Freidreich's ataxia, amylotrophic lateral scleorsis, hemophilia, Huntington's disease, retinal dystrophy, Rett syndrome, sickle cell disease, or a combination thereof.

[495] In some embodiments, the method of treating a disease comprises inserting a gene encoding a functional human protein into the target nucleic acid. In some embodiments, the target nucleic acid is recited in **TABLE 10**. In some embodiments, the gene comprises a nucleotide sequence encoding an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of the amino acid sequences recited in **TABLE 11**.

[496] In some embodiments, the method for treating is performed in a cell. In some embodiments, the method for treating is performed in vivo. In some embodiments, the cell is hepatocyte. In some embodiments, the cell is a mammalian cell. In some embodiments, the cell is a human cell. In some embodiments, the cell is a: stem cell, progenitor cell, induced pluripotent stem cell (iPSC) or a cell derived from an iPSC.

[497] In some embodiments, a Cas protein is used to effect the insertion. Cas proteins may be fused to transcription activators or transcriptional repressors or deaminases or other nucleic acid modifying proteins. In some embodiments, Cas proteins need not be fused to a partner protein to accomplish the required protein (expression) modification.

[498] In some embodiments, treatment of a disease comprises administration of a gene therapy. "Gene therapy", as used herein, comprises use of a recombinant nucleic acid (DNA or RNA), administered for the purpose to add a gene sequence. In some embodiments, a gene therapy comprises use of a vector to introduce a functional gene or transgene. In some embodiments, vectors comprise nonviral vectors, including cationic polymers, cationic lipids, or bio-responsive polymers. In some embodiments, the bio-responsive polymer exploits chemical-physical properties of the endosomal environment (*e.g.*, pH) to preferentially release the genetic material in the intracellular space. In some embodiments, vectors comprise viral vectors, including retroviruses, adenoviruses, adeno-associated

viruses, and herpes simplex viruses. In some embodiments, the vector comprises a replication-defective viral vector, comprising an insertion of a therapeutic gene inserted in genes essential to the lytic cycle, preventing the virus from replicating and exerting cytotoxic effects. Methods of gene therapy are described in more detail in Ingusci et al., "Gene Therapy Tools for Brain Diseases", Front. Pharmacol. 10:724 (2019) which is hereby incorporated by reference in its entirety.

[499] In some embodiments, a transgene is a nucleotide sequence that is inserted into a cell for expression of said nucleotide sequence in the cell. A transgene is meant to include (1) a nucleotide sequence that is not naturally found in the cell (*e.g.*, a heterologous nucleotide sequence); (2) a nucleotide sequence that is a mutant form of a nucleotide sequence naturally found in the cell into which it has been introduced; (3) a nucleotide sequence that serves to add additional copies of the same (*e.g.*, exogenous or homologous) or a similar nucleotide sequence naturally occurring in the cell into which it has been introduced; or (4) a silent naturally occurring or homologous nucleotide sequence whose expression is induced in the cell into which it has been introduced. The cell in which transgene expression occurs can be a target cell, such as a host cell.

[500] It is known that CRISPR-Cas9 gene editing techniques may select for p53-mutated cells. Similarly, the presence of KRAS mutations provides a selective advantage during CRISPR-Cas9 gene editing, as further described in Sinha et al., "A systematic genome-wide mapping of oncogenic mutation selection during CRISPR-Cas9 genome editing", Nature Comm. 12:6512 (2021), which is hereby incorporated by reference in its entirety. In some embodiments, a gene inserted for treatment comprises a nucleotide sequence that encodes wild-type functional human protein or a variant thereof. In some embodiments, the mutated donor nucleic acid encodes a function human protein.

[501] In some embodiments, treating, preventing, or inhibiting disease or disorder in a subject may comprise contacting a target nucleic acid associated with a particular ailment with a composition described herein. In some aspects, the methods of treating, preventing, or inhibiting a disease or disorder may involve removing, modifying, replacing, transposing, or affecting the regulation of a genomic sequence of a patient in need thereof. In some embodiments, the methods of treating, preventing, or inhibiting a disease or disorder may involve modulating gene expression.

[502] Described herein are compositions and methods for treating a disease in a subject by editing a target nucleic acid associated with a gene or expression of a gene related to the disease. In some embodiments, methods comprise administering a composition or cell described herein to a subject. By way of non-limiting example, the disease may be a cancer, an ophthalmological disorder, a neurological disorder, a neurodegenerative disease, a blood disorder, or a metabolic disorder, or a combination thereof. The disease may be an inherited disorder, also referred to as a genetic disorder. The disease may be the result of an infection or associated with an infection.

[503] In some embodiments, methods of treating a disease in a subject comprises administering to the subject the composition described herein, or the components of the systems described herein. In some embodiments, at least two components of the system are administered separately. In some

embodiments, at least two components of the system are administered simultaneously. In some embodiments, the subject treated by the methods described herein has any one of the diseases described herein. In some embodiments, administering to the subject the composition described herein, or the components of the systems described herein, increases the amount of the functional human protein in the subject (*e.g.*, in a biological sample from the subject) as compared to the subject before administration. Such an increase in the amount of the functional human protein can be by at least 10%, at least 20%, at least 30%, at least 40%, or at least 50% relative to the amount or concentration of the functional human protein in the subject (*e.g.*, in a biological sample from the subject) before administration. In some embodiments, the biological sample is selected from blood, serum, plasma, urine, saliva, and cerebrospinal fluid.

[504] The compositions and methods described herein may be used to treat, prevent, or inhibit a disease or syndrome in a subject. In some embodiments, the disease is a liver disease, a lung disease, an eye disease, or a muscle disease. Exemplary diseases and syndromes include but are not limited to the diseases and syndromes listed in **TABLE 12**.

[505] In some embodiments, compositions and methods edit at least one gene associated with a disease described herein or the expression thereof. In some embodiments, the disease is Alzheimer's disease and the gene is selected from *APP*, *BACE-1*, *PSD95*, *MAPT*, *PSEN1*, *PSEN2*, and *APOEε4*. In some embodiments, the disease is Parkinson's disease and the gene is selected from *SNCA*, *GDNF*, and *LRRK2*. In some embodiments, the disease comprises Centronuclear myopathy and the gene is *DNM2*. In some embodiments, the disease is Huntington's disease and the gene is *HTT*. In some embodiments, the disease is Alpha-1 antitrypsin deficiency (AATD) and the gene is *SERPINA1*. In some embodiments, the disease is amyotrophic lateral sclerosis (ALS) and the gene is selected from *SOD1*, *FUS*, *C9ORF72*, *ATXN2*, *TARDBP*, and *CHCHD10*. In some embodiments, the disease comprises Alexander Disease and the gene is *GFAP*. In some embodiments, the disease comprises Angelman Syndrome and the gene is *UBE3A*. In some embodiments, the disease comprises calcific aortic stenosis and the gene is Apo(a). In some embodiments, the disease comprises CD3Z-associated primary T-cell immunodeficiency and the gene is *CD3Z* or *CD247*. In some embodiments, the disease comprises CD18 deficiency and the gene is *ITGB2*. In some embodiments, the disease comprises CD40L deficiency and the gene is *CD40L*. In some embodiments, the disease is congenital adrenal hyperplasia and the gene is *CAH1*. In some embodiments, the disease comprises CNS trauma and the gene is *VEGF*. In some embodiments, the disease comprises coronary heart disease and the gene is selected from *FGA*, *FGB*, and *FGG*. In some embodiments, the disease comprises MECP2 Duplication syndrome and Rett syndrome and the gene is *MECP2*. In some embodiments, the disease comprises a bleeding disorder (coagulation) and the gene is FXI. In some embodiments, the disease comprises fragile X syndrome and the gene is *FMR1*. In some embodiments, the disease comprises Fuchs corneal dystrophy and the gene is selected from *ZEB1*, *SLC4A11*, and *LOXHD1*. In some embodiments, the disease comprises GM2-Gangliosidoses (*e.g.*, Tay Sachs Disease, Sandhoff disease) and the gene is selected from *HEXA* and *HEXB*. In some embodiments, the disease comprises Hearing loss disorders and the gene is *DFNA36*. In some embodiments, the disease is Pompe disease, including infantile onset Pompe disease

(IOPD) and late onset Pompe disease (LOPD) and the gene is *GAA*. In some embodiments, the disease is Retinitis pigmentosa and the gene is selected from *PDE6B, RHO, RP1, RP2, RPGR, PRPH2, IMPDH1, PRPF31, CRB1, PRPF8, TULP1, CA4, HPRPF3, ABCA4, EYS, CERKL, FSCN2, TOPORS, SNRNP200, PRCD, NR2E3, MERTK, USH2A, PROM1, KLHL7, CNGB1, TTC8, ARL6, DHDDS, BEST1, LRAT, SPARA7, CRX, CLRN1, RPE65,* and *WDR19*. In some embodiments, the disease comprises Leber Congenital Amaurosis Type 10 and the gene is *CEP290*. In some embodiments, the disease is cardiovascular disease and/or lipodystrophies and the gene is selected from *ABCG5, ABCG8, AGT, ANGPTL3, APOCIII, APOA1, APOL1, ARH, CDKN2B, CFB, CXCL12, FXI, FXII, GATA-4, MIA3, MKL2, MTHFD1L, MYH7, NKX2-5, NOTCH1, PKK, PCSK9, PSRC1, SMAD3,* and *TTR*. In some embodiments, the disease comprises acromegaly and the gene is *GHR*. In some embodiments, the disease is diabetes and the gene is *GCGR*. In some embodiments, the disease is NAFLD/NASH and the gene is selected from *DGAT2* and *PNPLA3*. In some embodiments, the disease is cystic fibrosis and the gene is *CFTR*. In some embodiments, the disease is Duchenne muscular dystrophy and the gene is *DMD*. In some embodiments, the disease is ornithine transcarbamylase deficiency and the gene is *OTC*. In some embodiments, the disease comprises angioedema and the gene is *PKK*. In some embodiments, the disease comprises thalassemia and the gene is *TMPRSS6*. In some embodiments, the disease comprises achondroplasia and the gene is *FGFR3*. In some embodiments, the disease comprises Cri du chat syndrome and the gene is selected from *CTNND2*. In some embodiments, the disease comprises sickle cell anemia and the gene is Beta globin gene. In some embodiments, the disease comprises Alagille Syndrome and the gene is selected from *JAG1* and *NOTCH2*. In some embodiments, the disease comprises Charcot-Marie-Tooth disease and the gene is selected from *PMP22* and *MFN2*. In some embodiments, the disease comprises Crouzon syndrome and the gene is selected from *FGFR2*, *FGFR3*, and *FGFR3*. In some embodiments, the disease comprises Dravet Syndrome and the gene is selected from *SCN1A* and *SCN2A*. In some embodiments, the disease comprises Emery-Dreifuss syndrome and the gene is selected from *EMD, LMNA, SYNE1, SYNE2, FHL1*, and *TMEM43*. In some embodiments, the disease comprises Factor V Leiden thrombophilia and the gene is *F5*. In some embodiments, the disease is fabry disease and the gene is *GLA*. In some embodiments, the disease is facioscapulohumeral muscular dystrophy and the gene is *FSHD1*. In some embodiments, the disease comprises Fanconi anemia and the gene is selected from *FANCA, FANCB, FANCC, FANCD1, FANCD2, FANCE, FANCF, FANCG, FANCI, FANCJ, FANCL, FANCM, FANCN, FANCP, FANCS, RAD51C*, and *XPF*. In some embodiments, the disease comprises Familial Creutzfeld–Jakob disease and the gene is *PRNP*. In some embodiments, the disease comprises Familial Mediterranean Fever and the gene is *MEFV*. In some embodiments, the disease comprises Friedreich's ataxia and the gene is *FXN*. In some embodiments, the disease comprises Gaucher disease and the gene is *GBA*. In some embodiments, the disease comprises hemochromatosis and the gene is *HFE*, optionally comprising a C282Y mutation. In some embodiments, the disease comprises Hemophilia A and the gene is *FVIII*. In some embodiments, the disease is hereditary angioedema and the gene is *SERPING1*. In some embodiments, the disease

comprises histiocytosis and the gene is *CD1*. In some embodiments, the disease comprises immunodeficiency 17 and the gene is *CD3D*. In some embodiments, the disease comprises immunodeficiency 13 and the gene is *CD4*. In some embodiments, the disease comprises Common Variable Immunodeficiency and the gene is selected from *CD19* and *CD81*. In some embodiments, the disease comprises Joubert syndrome and the gene is selected from *INPP5E, TMEM216, AHI1, NPHP1, CEP290, TMEM67, RPGRIP1L, ARL13B, CC2D2A, OFD1, TMEM138, TCTN3, ZNF423,* and *AMRC9*. In some embodiments, the disease comprises leukocyte adhesion deficiency and the gene is *CD18*. In some embodiments, the disease comprises Li–Fraumeni syndrome and the gene is *TP53*. In some embodiments, the disease comprises Lynch syndrome and the gene is selected from *MSH2, MLH1, MSH6, PMS2, PMS1, TGFBR2,* and *MLH3*. In some embodiments, the disease comprises Marfan syndrome and the gene is *FBN1*. In some embodiments, the disease comprises mastocytosis and the gene is *CD2*. In some embodiments, the disease comprises methylmalonic acidemia and the gene is selected from *MMAA*, *MMAB*, and *MUT*. In some embodiments, the disease is myotonic dystrophy and the gene is selected from *CNBP* and *DMPK*. In some embodiments, the disease comprises neurofibromatosis and the gene is selected from *NF1*, and *NF2*. In some embodiments, the disease comprises osteogenesis imperfecta and the gene is selected from *COL1A1*, *COL1A2*, and *IFITM5*. In some embodiments, the disease comprises Peutz–Jeghers syndrome and the gene is *STK11*. In some embodiments, the disease comprises polycystic kidney disease and the gene is selected from *PKD1* and *PKD2*. In some embodiments, the disease comprises Severe Combined Immune Deficiency and the gene is selected from *IL7R, RAG1,* and *JAK3*. In some embodiments, the disease comprises PRKAG2 cardiac syndrome and the gene is *PRKAG2*. In some embodiments, the disease comprises spinocerebellar ataxia and the gene is selected from *ATXN1, ATXN2, ATXN3, PLEKHG4, SPTBN2, CACNA1A, ATXN7, ATXN8OS, ATXN10, TTBK2, PPP2R2B, KCNC3, PRKCG, ITPR1, TBP, KCND3,* and *FGF14*. In some embodiments, the disease is thrombophilia due to antithrombin III deficiency and the gene is *SERPINC1*. In some embodiments the disease is spinal muscular atrophy and the gene is *SMN1*. In some embodiments, the disease comprises Usher Syndrome and the gene is selected from *MYO7A, USH1C, CDH23, PCDH15, USH1G, USH2A, GPR98, DFNB31,* and *CLRN1*. In some embodiments, the disease comprises von Willebrand disease and the gene is *VWF*. In some embodiments, the disease comprises Waardenburg syndrome and the gene is selected from *PAX3, MITF, WS2B, WS2C, SNAI2, EDNRB, EDN3,* and *SOX10*. In some embodiments, the disease comprises Wiskott-Aldrich Syndrome and the gene is *WAS*. In some embodiments, the disease comprises von Hippel–Lindau disease and the gene is *VHL*. In some embodiments, the disease comprises Wilson disease and the gene is *ATP7B*. In some embodiments, the disease comprises Zellweger syndrome and the gene is selected from *PEX1, PEX2, PEX3, PEX5, PEX6, PEX10, PEX12, PEX13, PEX14, PEX16, PEX19,* and *PEX26*. In some embodiments, the disease comprises infantile myofibromatosis and the gene is *CD34*. In some embodiments, the disease comprises platelet glycoprotein IV deficiency and the gene is *CD36*. In some embodiments, the disease comprises immunodeficiency with hyper-IgM type 3

and the gene is *CD40*. In some embodiments, the disease comprises hemolytic uremic syndrome and the gene is *CD46*. In some embodiments, the disease comprises complement hyperactivation, angiopathic thrombosis, or protein-losing enteropathy and the gene is *CD55*. In some embodiments, the disease comprises hemolytic anemia and the gene is *CD59*. In some embodiments, the disease comprises calcification of joints and arteries and the gene is *CD73*. In some embodiments, the disease comprises immunoglobulin alpha deficiency and the gene is *CD79A*. In some embodiments, the disease comprises C syndrome and the gene is *CD96*. In some embodiments, the disease comprises histiocytic sarcoma and the gene is *CD163*. In some embodiments, the disease comprises autosomal dominant deafness and the gene is *CD164*. In some embodiments, the disease comprises immunodeficiency 25 and the gene is *CD247*. In some embodiments, the disease comprises methymalonic acidemia due to transcobalamin receptor defect and the gene is *CD320*.

XII. Illustrative Embodiments

[506] Embodiment 1. A composition comprising:

i. an effector protein, or a nucleic acid encoding the effector protein;

ii. a guide nucleic acid, or a nucleic acid encoding the guide nucleic acid; and

iii. a donor nucleic acid,

wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**,

wherein the guide nucleic acid comprises a spacer sequence that hybridizes to a target sequence in a target nucleic acid,

wherein the target sequence comprises a nucleotide sequence within a human safe harbor locus, and

wherein the donor nucleic acid encodes a transgene that comprises a functional human protein that is expressed upon incorporation into the human safe harbor locus.

[507] Embodiment 2. The composition of embodiment 1, wherein the effector protein comprises an amino acid sequence with at least 95% sequence identity to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**.

[508] Embodiment 3. The composition of embodiment 1 or 2, wherein the effector protein comprises one or more amino acid substitutions relative to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**.

[509] Embodiment 4. The composition of embodiment 3, wherein the one or more amino acid substitutions independently comprise one or more conservative substitutions, one or more non-conservative substitutions, or combinations thereof.

[510] Embodiment 5. The composition of embodiment 3 or 4, wherein the one or more amino acid substitutions comprise one or more substitutions with a positively charged amino acid residues.

[511] Embodiment 6. The composition of embodiment 5, wherein the positively charged amino acid residue is independently selected from Lys (K), Arg (R), and His (H).

[512] Embodiment 7. The composition of any one of embodiments 1-6, wherein the target sequence is in proximity to or adjacent to a protospacer adjacent motif (PAM) sequence comprising any one of the amino acid sequences recited in **TABLE 3**.

[513] Embodiment 8. The composition of any one of embodiments 1-7, wherein the effector protein comprises a nuclear localization signal.

[514] Embodiment 9. The composition of any one of embodiments 1-8, comprising a fusion partner protein linked to the effector protein.

[515] Embodiment 10. The composition of embodiment 9, wherein the fusion partner protein is directly fused to the N terminus or C terminus of the effector protein via an amide bond.

[516] Embodiment 11. The composition of any one of embodiments 1-10, wherein the nuclear localization signal comprises an amino acid sequence comprising any one of the amino acid sequences recited in **TABLE 2.**

[517] Embodiment 12. The composition of any one of embodiments 1-11, wherein the human safe harbor locus is an intron.

[518] Embodiment 13. The composition of any one of embodiments 1-11, wherein the human safe harbor locus is a gene or portion thereof that is expressed in the liver.

[519] Embodiment 14. The composition of any one of embodiments 1-13, wherein the human safe harbor locus comprises any one of the nucleotide sequences recited **TABLE 10**.

[520] Embodiment 15. The composition of any one of embodiments 1-13, wherein the human safe harbor locus is located in human chromosome 1, human chromosome 3, human chromosome 4, human chromosome 6, human chromosome 10, human chromosome 11, human chromosome 12, human chromosome 14, human chromosome 17, human chromosome 18, or human chromosome 19.

[521] Embodiment 16. The composition of any one of embodiments 1-15, wherein the guide nucleic acid comprises a nucleotide sequence that is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 4**, **TABLE 5**, **TABLE 6**, **TABLE 7**, **TABLE 9,** and any combination thereof.

[522] Embodiment 17. The composition of any one of embodiments 1-16, wherein the guide nucleic acid comprises a spacer sequence that is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 4** and **TABLE 5**.

[523] Embodiment 18. The composition of any one of embodiments 1-17, wherein the guide nucleic acid does not comprise a tracrRNA.

[524] Embodiment 19. The composition of any one of embodiments 1-18, wherein the target sequence is a human gene.

[525] Embodiment 20. The composition of any one of embodiments 1-19, wherein the target sequence comprises a nucleotide sequence with at least 90% sequence identity to any one of nucleotide sequences recited in **TABLE 10**.

[526] Embodiment 21. The composition of any one of embodiments 1-20, wherein the composition results in cleavage of both strands of the human safe harbor locus prior to incorporation of the donor nucleic acid.

[527] Embodiment 22. The composition of any one of embodiments 1-21, wherein the functional human protein is select from any one of CFTR, DMD, GAA, A1AT, FXN, F8, F9, SOD1, C9, HTT, MECP2, SMN1, TARDBP, FUS, RHO, and USH2A or a functional variant or fragment thereof.

[528] Embodiment 23. The composition of any one of embodiments 1-22, wherein the target sequence comprises a nucleotide sequence within **SEQ ID NO: 266**.

[529] Embodiment 24. The composition of embodiment 23, wherein the spacer sequence is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 4**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6, 13, 47** and **228-231**.

[530] Embodiment 25. The composition of embodiment 24, wherein the guide nucleic acid comprises a spacer sequence that is at least 90% identical to any one of **SEQ ID NO: 348, 558, 401, 409, 473** and **486**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6** and **228-230**.

[531] Embodiment 26. The composition of embodiment 24, wherein the guide nucleic acid comprises a spacer sequence that is at least 90% identical to any one of **SEQ ID NO: 592, 600** and **602**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 231**.

[532] Embodiment 27. The composition of embodiment 24, wherein the guide nucleic acid comprises a nucleotide sequence that is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 9**.

[533] Embodiment 28. The composition of embodiment 23, wherein the guide nucleic acid comprises a nucleotide sequence that is at least 90% identical to any one of **SEQ ID NO: 652, 685, 705, 713, 777** and **790**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6** and **228-230**.

[534] Embodiment 29. The composition of embodiment 27, wherein the guide nucleic acid comprises a nucleotide sequence that is at least 90% identical to any one of **SEQ ID NO: 856, 864** and **866**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 231**.

[535] Embodiment 30. The composition of any one of embodiments 1-22, wherein the target sequence comprises a nucleotide sequence within **SEQ ID NO: 1795**.

[536] Embodiment 31. The composition of embodiment 30, wherein the spacer sequence is at least 90% identical to any one of nucleotide sequences recited in **TABLE 5**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of amino acid sequences of **SEQ ID NO: 6, 13, 47,** and **228-231**.

[537] Embodiment 32. A composition comprising:

i. an effector protein, or a nucleic acid encoding the effector protein;

ii. a guide nucleic acid, or a nucleic acid encoding the guide nucleic acid; and

iii. a donor nucleic acid,

wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**,

wherein the guide nucleic acid is a crRNA,

wherein the crRNA comprises a spacer sequence and a repeat sequence,

wherein the spacer sequence hybridizes to a target sequence in a target nucleic acid,

wherein the target sequence comprises a nucleotide sequence within a human safe harbor locus,

wherein the target sequence comprises any one of the nucleotide sequences recited in **TABLE 10**, and

wherein the donor nucleic acid encodes a functional human protein that upon expression after introduction in a human subject expression replaces the function of a corresponding nonworking or missing gene in the human subject.

[538] Embodiment 33. The composition of embodiment 32, wherein the effector protein comprises one or more amino acid substitutions relative to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**.

[539] Embodiment 34. The composition of embodiment 33, wherein the one or more amino acid substitutions independently comprise one or more conservative substitutions, one or more non-conservative substitutions, or combinations thereof.

[540] Embodiment 35. The composition of embodiment 33 or 34, wherein the one or more amino acid substitutions comprise one or more substitutions with a positively charged amino acid residues.

[541] Embodiment 36. The composition of embodiment 35, wherein the positively charged amino acid residue is independently selected from Lys (K), Arg (R), and His (H).

[542] Embodiment 37. The composition of any one of embodiments 32-36, wherein the functional human protein is selected from any one of A1AT, CFTR, DMD, FXN, F8, F9, GAA, SOD1, C9, HTT, MECP2, SMN1, TARDBP, FUS, RHO, and USH2A or a functional variant or fragment thereof.

[543] Embodiment 38. The composition of embodiment 37, wherein the composition is formulated for administration to a human subject suffering from a disorder characterized by irregular expression of the functional human protein absent the administration.

[544] Embodiment 39. The composition of any one of embodiments 36-38, wherein the spacer sequence is at least 90% identical to any one of nucleotide sequences recited in **TABLE 4** and **TABLE 5**.

[545] Embodiment 40. The composition of any one of embodiments 36-38, wherein the repeat sequence that is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 6**.

[546] Embodiment 41. The composition of any one of embodiments 36-39, wherein the repeat sequence is identical to any one of **SEQ ID NO: 252-258, 1789,** and **1848**.

[547] Embodiment 42. The composition of any one of embodiments 36-40, wherein the target sequence comprises a nucleotide sequence within **SEQ ID NO: 266**.

[548] Embodiment 43. The composition of embodiment 42, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of amino acid sequences of **SEQ ID NO: 6, 47** and **228-230.**

[549] Embodiment 44. The composition of embodiment 42 or 43, wherein the spacer sequence is at least 90% identical to any one of **SEQ ID NO: 301-510** and **536-575**.

[550] Embodiment 45. The composition of embodiment 44, wherein the spacer sequence is at least 90% identical to any one of **SEQ ID NO: 348, 401, 409, 473, 486,** and **558**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6,** and **228-230**.

[551] Embodiment 46. The composition of embodiment 45, wherein the crRNA comprises a nucleotide sequence that is at least 90% identical to any one of **SEQ ID NO: 605-839**.

[552] Embodiment 47. The composition of embodiment 46, wherein the crRNA comprises a nucleotide sequence that is at least 90% identical to any one of **SEQ ID NO: 652, 685, 705, 713, 777** and **790**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6** and **228-230**.

[553] Embodiment 48. The composition of any one of embodiments 32-47, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of the amino acid sequences of **SEQ ID NO: 6** and **228-230**, and wherein the effector protein comprises one or more amino acid substitutions independently selected from E109R, H208R, K184R, K38R, L182R, Q183R, S108R, S198R, T114R or a combination thereof.

[554] Embodiment 49. A composition comprising:

i. an effector protein, or a nucleic acid encoding the effector protein;

ii. a guide nucleic acid, or a nucleic acid encoding the guide nucleic acid; and

iii. a donor nucleic acid,

wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**,

wherein the guide nucleic acid is a single guide RNA (sgRNA),

wherein the sgRNA comprises a spacer sequence and a handle sequence,

wherein the spacer sequence hybridizes to a target sequence in a target nucleic acid,

wherein the handle sequence comprises one or more of an intermediary sequence, a repeat sequence, and a linker,

wherein the target sequence comprises a nucleotide sequence within a human safe harbor locus,

wherein the target sequence comprises any one of the nucleotide sequences recited in **TABLE 10,** and

wherein the donor nucleic acid encodes a functional human protein that upon expression after introduction in a human subject expression replaces the function of a corresponding nonworking or missing gene in the human subject.

[555] Embodiment 50. The composition of embodiment 49, wherein the effector protein comprises one or more amino acid substitutions relative to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**.

[556] Embodiment 51. The composition of embodiment 50, wherein the one or more amino acid substitutions independently comprise one or more conservative substitutions, one or more non-conservative substitutions, or combinations thereof.

[557] Embodiment 52. The composition of embodiment 50 or 51, wherein the one or more amino acid substitutions comprise one or more substitutions with a positively charged amino acid residues.

[558] Embodiment 53. The composition of embodiment 52, wherein the positively charged amino acid residue is independently selected from Lys (K), Arg (R), and His (H).

[559] Embodiment 54. The composition of any one of embodiments 49-53, wherein the spacer sequence is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 4** and **TABLE 5**.

[560] Embodiment 55. The composition of any one of embodiments 49-54, wherein the repeat sequence that is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 6.**

[561] Embodiment 56. The composition of any one of embodiments 49-55, wherein the repeat sequence is identical to any one of **SEQ ID NO: 252-258, 1789** and **1848**.

[562] Embodiment 57. The composition of any one of embodiments 49-56, wherein the handle is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 7.**

[563] Embodiment 58. The composition of any one of embodiments 49-57, wherein the handle sequence is identical in a range of from 90% to 100% to any one of **SEQ ID NO: 259-264**.

[564] Embodiment 59. The composition of any one of embodiments 49-58, wherein the linker sequence comprises a nucleotide sequence recited in **TABLE 8**.

[565] Embodiment 60. The composition of any one of embodiments 49-59, wherein the linker sequence comprises a nucleotide sequence of **SEQ ID NO: 265.**

[566] Embodiment 61. The composition of any one of embodiments 49-60, wherein the target sequence comprises a nucleotide sequence within **SEQ ID NO: 266**.

[567] Embodiment 62. The composition of any one of embodiments 49-61, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 13** and **231.**

[568] Embodiment 63. The composition of any one of embodiments 49-62, wherein the spacer sequence is at least 90% identical to any one of **SEQ ID NO: 536-604**.

[569] Embodiment 64. The composition of embodiment 63, wherein the spacer sequence is at least 90% identical to any one of **SEQ ID NO: 592, 600** and **602**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 231**.

[570] Embodiment 65. The composition of any one of embodiments 49-63, wherein the sgRNA comprises a nucleotide sequence that is at least 90% identical to any one of **SEQ ID NO: 840-908**.

[571] Embodiment 66. The composition of embodiment 65, wherein the sgRNA comprises a nucleotide sequence that is at least 90% identical to any one of **SEQ ID NO: 856, 864** and **866**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 231**.

[572] Embodiment 67. The composition of any one of embodiments 49-66, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 231**, and wherein the effector protein comprises one or more amino acid substitutions independently selected from K58W, I80K, N193K, S209F, A218K, E225K, N286K, M295W, M298L, A306K, Y315M or a combination thereof.

[573] Embodiment 68. The composition of any one of embodiments 49-66, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 231**, and wherein the effector protein comprises one or more amino acid substitutions independently selected from I80R, T84R, K105R, G210R, C202R, A218R, D220R, E225R, C246R, Q360R or a combination thereof.

[574] Embodiment 69. A composition comprising:

i. an effector protein, or a nucleic acid encoding the effector protein;

ii. a guide nucleic acid, or a nucleic acid encoding the guide nucleic acid; and

iii. a donor nucleic acid,

wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of the amino acid sequences recited in **SEQ ID NO: 47**,

wherein the guide nucleic acid comprises a spacer sequence that hybridizes to a target sequence in a target nucleic acid,

wherein the target sequence comprises a nucleotide sequence within intron 1 of a mammalian albumin gene, and

wherein the donor nucleic acid encodes a functional human protein that upon expression after introduction in a human subject expression replaces the function of a corresponding nonworking or missing gene in the human subject.

[575] Embodiment 70. The composition of embodiment 69, wherein the effector protein comprises an amino acid sequence with at least 95% sequence identity to the amino acid sequence of **SEQ ID NO: 47**.

[576] Embodiment 71. The composition of embodiment 69 or 70, wherein the effector protein recognizes a protospacer adjacent motif (PAM) sequence comprising any one of the nucleotide sequences recited in **TABLE 3**.

[577] Embodiment 72. The composition of any one of embodiments 69-71, wherein the effector protein comprises a nuclear localization signal.

[578] Embodiment 73. The composition of embodiment 72, wherein the nuclear localization signal comprises an amino acid sequence that is identical to an amino acid sequence of **TABLE 2**.

[579] Embodiment 74. The composition of any one of embodiments 69-73, further comprising a fusion partner protein linked to the effector protein.

[580] Embodiment 75. The composition of embodiment 74, wherein the fusion partner protein is directly fused to the N terminus or C terminus of the effector protein via an amide bond.

[581] Embodiment 76. The composition of any one of embodiments 69-75, wherein the spacer sequence comprises a nucleotide sequence that has at least 90% sequence identity to any one of nucleotide sequences recited in **TABLE 4**.

[582] Embodiment 77. The composition of any one of embodiments 69-76, wherein the spacer sequence comprises a nucleotide sequence that is at least 90% identical to a nucleotide sequence that hybridizes to a target sequence having at least 17 contiguous nucleotides of any one of the nucleotide sequences recited in **TABLE 10** or a reverse complement thereof.

[583] Embodiment 78. The composition of any one of embodiments 69-77, wherein the guide nucleic acid comprises a repeat sequence that is at least 90% identical to **SEQ ID NO: 254**.

[584] Embodiment 79. The composition of any one of embodiments 69-78, wherein the guide nucleic acid comprises a nucleotide sequence that is at least 90% identical to any one of the guide RNA sequences recited in **TABLE 9**.

[585] Embodiment 80. The composition of any one of embodiments 69-79, wherein the guide nucleic acid does not comprise a tracrRNA.

[586] Embodiment 81. The composition of any one of embodiments 69-80, wherein the guide nucleic acid is a crRNA.

[587] Embodiment 82. The composition of any one of embodiments 69-81, wherein the target sequence is a human gene.

[588] Embodiment 83. The composition of any one of embodiments 69-82, wherein the target sequence comprises at least 17 contiguous nucleotides of any one of the nucleotide sequences recited in **TABLE 10** or a reverse complement thereof.

[589] Embodiment 84. The composition of any one of embodiments 69-83, wherein the composition cleaves both strands of the human albumin gene.

[590] Embodiment 85. The composition of any one of embodiments 69-84, wherein the donor nucleic acid comprises at least 90% sequence identity to any one of the nucleotide sequences recited in **TABLE 11**.

[591] Embodiment 86. A composition comprising:

an effector protein, or a nucleic acid encoding the effector protein;

a guide nucleic acid, or a nucleic acid encoding the guide nucleic acid; and

a donor nucleic acid,

wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 47**,

wherein the guide nucleic acid is a crRNA,

wherein the crRNA comprises a spacer sequence and a repeat sequence,

wherein the spacer sequence hybridizes to a target sequence in a target nucleic acid,

wherein the spacer sequence is identical to any one of **SEQ ID NO: 551-575**,

wherein the repeat sequence is identical to **SEQ ID NO: 254**,

wherein the target sequence comprises a nucleotide sequence within **SEQ ID NO: 266**, and

wherein the donor nucleic acid encodes a functional human protein that upon expression after introduction in a human subject expression replaces the function of a corresponding nonworking or missing gene in the human subject.

[592] Embodiment 87. The composition of embodiment 86, wherein the guide nucleic acid is at least 90% identical to any one of **SEQ ID NO: 815-839**.

[593] Embodiment 88. The composition of embodiment 86, wherein the crRNA is at least 90% identical to any one of **SEQ ID NO: 551-575**.

[594] Embodiment 89. The composition of any one of embodiments 1-87, comprising a nucleic acid expression vector, wherein the expression vector comprises at least one of the nucleic acid encoding the effector protein; the nucleic acid encoding the guide nucleic acid; and the donor nucleic acid.

[595] Embodiment 90. The composition of embodiment 89, wherein the nucleic acid expression vector is a viral vector.

[596] Embodiment 91. The composition of embodiment 90, wherein the viral vector is an adeno associated viral (AAV) vector.

[597] Embodiment 92. The composition of any one of embodiments 1-91, wherein the nucleic acid encoding the effector protein is a messenger RNA.

[598] Embodiment 93. The composition of any one of embodiments 1-92, comprising a lipid or a lipid nanoparticle.

[599] Embodiment 94. A pharmaceutical composition, comprising the composition of any one of embodiments 1-93; and a pharmaceutically acceptable excipient.

[600] Embodiment 95. A system comprising components for introduction of a donor nucleic acid encoding a functional human protein into a human safe harbor locus, wherein the components comprise:

an effector protein, or a nucleic acid encoding the effector protein, wherein the effector protein comprises an amino acid sequence having at least 90% sequence identity to any one of amino acid sequences recited in **TABLE 1A** and **TABLE 1B**;

a guide nucleic acid, or a nucleic acid encoding the guide nucleic acid, wherein the guide nucleic acid comprises a spacer sequence that hybridizes to a target sequence in a target nucleic acid, and wherein the target sequence comprises a nucleotide sequence within the human safe harbor locus; and

the donor nucleic acid.

[601] Embodiment 96. The system of embodiment 95, wherein the effector protein comprises one or more amino acid substitutions relative to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**.

[602] Embodiment 97. The system of embodiment 96, wherein the one or more amino acid substitutions independently comprise one or more conservative substitutions, one or more non-conservative substitutions, or combinations thereof.

[603] Embodiment 98. The system of embodiment 96 or 97, wherein the one or more amino acid substitutions comprise one or more substitutions with a positively charged amino acid residues.

[604] Embodiment 99. The system of embodiment 98, wherein the positively charged amino acid residue is independently selected from Lys (K), Arg (R), and His (H).

[605] Embodiment 100. The system of any one of embodiments 95-99, wherein the target sequence is in proximity to or adjacent to a protospacer adjacent motif (PAM) sequence comprising any one of the nucleotide sequences recited in **TABLE 3**, and wherein the effector protein and guide form a complex that modifies the human safe harbor locus upon hybridization of the guide nucleic acid to the target sequence and recognition of the PAM sequence by the effector protein.

[606] Embodiment 101. The system of any one of embodiments 95-100, wherein the target sequence comprises any one of the nucleotide sequences recited in **TABLE 10**.

[607] Embodiment 102. The system of embodiment 101, wherein the spacer sequence is at least 90% identical to any one of nucleotide sequences recited in **TABLE 4** and **TABLE 5**, and wherein the

effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6, 13, 47** and **228-231**.

[608] Embodiment 103. The system of any one of embodiment 95-102, wherein the target sequence comprises a nucleotide sequence within **SEQ ID NO: 266**.

[609] Embodiment 104. The system of embodiment 103, wherein the spacer sequence is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 4**.

[610] Embodiment 105. The system of embodiment 104, wherein the guide nucleic acid comprises a nucleotide sequence that is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 9**.

[611] Embodiment 106. A method of expressing a functional human protein in a cell, the method comprising contacting the cell with the composition of any one of embodiments 1-93, or the system of any one of embodiments 95-105.

[612] Embodiment 107. A method of treating a disease associated with a mutation or aberrant expression of a human protein in a subject in need thereof, the method comprising administering to the subject the composition of any one of embodiments 1-93.

[613] Embodiment 108. The method of embodiment 107, wherein the subject has a genetic disorder.

[614] Embodiment 109. The method of embodiment 108, wherein the genetic disorder is a monogenic disorder.

[615] Embodiment 110. The method of any one of embodiments 107-109, wherein the subject has cystic fibrosis, muscular dystrophy, Freidreich's ataxia, amylotrophic lateral scleorsis, hemophilia, Huntington's disease, retinal dystrophy, Rett syndrome, sickle cell disease, or a combination thereof.

[616] Embodiment 111. The method of any one of embodiments 107-110, wherein the subject has a reduced activity of the human protein prior to the administering.

[617] Embodiment 112. The method of any one of embodiments 107-110, wherein the subject has no activity of the human protein prior to the administering.

[618] Embodiment 113. The method of any one of embodiments 107-112, wherein the subject has one or more genetic mutations.

[619] Embodiment 114. The method of embodiment 113, wherein the one or more mutations comprise a point mutation, a single nucleotide polymorphism (SNP), a chromosomal mutation, a copy number mutation, or any combination thereof.

[620] Embodiment 115. The method of any one of embodiments 107-114, wherein the mutation is associated with one or more of protein expression, protein activity, and protein stability.

[621] Embodiment 116. The method of embodiment 115, wherein the mutation results in reduced expression of the human protein, reduced activity of human protein, reduced half-life of the human protein, or combinations thereof, in the cell relative to a cell without the mutation, before the treatment.

[622] Embodiment 117. A system comprising components for introduction of a donor nucleic acid into intron 1 of a human albumin gene, wherein the components comprise:

an effector protein, or a nucleic acid encoding the effector protein, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to the amino acid sequence recited in **SEQ ID NO: 47**;

a guide nucleic acid, or a nucleic acid encoding the guide nucleic acid, wherein the guide nucleic acid comprises a spacer sequence that hybridizes to a target sequence in a target nucleic acid, and wherein the target sequence comprises a nucleotide sequence within intron 1 of albumin gene; and

the donor nucleic acid, wherein the donor nucleic acid encodes a functional human protein that upon expression after introduction in a human subject expression replaces the function of a corresponding nonworking or missing gene in the human subject.

[623] Embodiment 118. The system of embodiment 117, wherein the guide nucleic acid comprises at least one nucleotide sequence selected from any one of **TABLES 3, 4,** and **5**.

[624] Embodiment 119. A method of expressing a functional human protein in a cell, the method comprising contacting the cell with the composition of any one of embodiments 69-93, or the system of embodiment 117 or 118.

[625] Embodiment 120. A method of treating a disease associated with a mutation or aberrant expression of a protein in a subject in need thereof, the method comprising administering to the subject the composition of any one of embodiments 69-93.

[626] Embodiment 121. The method of embodiment 120, wherein the subject has a genetic disorder.

[627] Embodiment 122. The method of embodiment 120 or 121, wherein the mutation comprises a point mutation, a single nucleotide polymorphism (SNP), a chromosomal mutation, a copy number mutation, or any combination thereof.

[628] Embodiment 123. The method of any one of embodiments 170-116 or 119-122, wherein the method is performed in a cell.

[629] Embodiment 124. The method of any one of embodiments 170-117 or 119-122, wherein the method is performed in vivo.

[630] Embodiment 125. A cell comprising the composition of any one of embodiments 1-93 or 69-87.

[631] Embodiment 126. A cell that comprises the target nucleic acid modified by the composition of any one of embodiments 1-93 or 69-87.

[632] Embodiment 127. The cell of any embodiment 125 or 126, wherein the cell is a hepatocyte.

[633] Embodiment 128. The cell of any one of embodiments 125-127, wherein the cell is a mammalian cell.

[634] Embodiment 129. The cell of any one of embodiments 125-128, wherein the cell is a human cell.

[635] Embodiment 130. The cell of any one of embodiments 125-129, wherein the cell is a: stem cell, progenitor cell, induced pluripotent stem cell (iPSC) or a cell derived from an iPSC.

[636] Embodiment 131. A population of cells that comprises at least one cell of any one of embodiments 125-130.

[637] Embodiment 132. A system for introduction of a donor nucleic acid into a human safe harbor comprising one or more components, wherein the one or more components individually comprises one or more of the following:

i. an effector protein, or a nucleic acid encoding the effector protein, wherein the effector protein comprises an amino acid sequence having at least 90% sequence identity to any one of amino acid sequences recited in **TABLE 1A** and **TABLE 1B**;

ii. a guide nucleic acid, or a nucleic acid encoding the guide nucleic acid, wherein the guide nucleic acid comprises a spacer sequence that hybridizes to a target sequence in a target nucleic acid, and wherein the target sequence comprises a nucleotide sequence within the human safe harbor; and

iii. the donor nucleic acid encoding a transgene that comprises a functional human protein that is expressed upon incorporation into the human safe harbor locus,

wherein the human safe harbor comprises at least 90% sequence identity to any one of the nucleotide sequences recited in **TABLE 10**.

[638] Embodiment 133. The system of embodiment 132, wherein the effector protein comprises an amino acid sequence with at least 95% sequence identity to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**.

[639] Embodiment 134. The system of embodiment 132 or 133, wherein the effector protein comprises one or more amino acid substitutions relative to any one of the sequences recited in **TABLE 1A** and **TABLE 1B**.

[640] Embodiment 135. The system of embodiment 134, wherein the one or more amino acid substitutions independently comprise one or more conservative substitutions, one or more non-conservative substitutions, or combinations thereof.

[641] Embodiment 136. The system of embodiment 134 or 135, wherein the one or more amino acid substitutions comprise one or more substitutions with a positively charged amino acid residues.

[642] Embodiment 137. The system of embodiment 136, wherein the positively charged amino acid residue is independently selected from Lys (K), Arg (R), and His (H).

[643] Embodiment 138. The system of any one of embodiments 132-137, wherein the target sequence is in proximity to or adjacent to a protospacer adjacent motif (PAM) sequence comprising any one of the nucleotide sequences recited in **TABLE 3**.

[644] Embodiment 139. The system of any one of embodiments 132-138, wherein the effector protein comprises a nuclear localization signal.

[645] Embodiment 140. The system of any one of embodiments 132-139, comprising a fusion partner protein linked to the effector protein.

[646] Embodiment 141. The system of embodiment 140, wherein the fusion partner protein is directly fused to the N terminus or C terminus of the effector protein via an amide bond.

[647] Embodiment 142. The system of any one of embodiments 132-141, wherein the human safe harbor locus is located in human chromosome 1, human chromosome 3, human chromosome 4, human chromosome 6, human chromosome 10, human chromosome 11, human chromosome 12, human chromosome 14, human chromosome 17, human chromosome 18, or human chromosome 19.

[648] Embodiment 143. The system of any one of embodiments 132-142, wherein the guide nucleic acid is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 4**, **TABLE 5**, **TABLE 6**, **TABLE 7**, **TABLE 9** and any combination thereof.

[649] Embodiment 144. The system of any one of embodiments 132-143, wherein the guide nucleic acid comprises a spacer sequence that is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 4** and **TABLE 5**.

[650] Embodiment 145. The system of any one of embodiments 132-144, wherein the guide nucleic acid does not comprise a tracrRNA.

[651] Embodiment 146. The system of any one of embodiments 132-145, wherein the system is capable of cleaving both strands of the human safe harbor locus prior to incorporation of the donor nucleic acid.

[652] Embodiment 147. The system of any one of embodiment 132-146, wherein the functional human protein is select from any one of CFTR, DMD, GAA, A1AT, FXN, F8, F9, SOD1, C9, HTT, MECP2, SMN1, TARDBP, FUS, RHO, and USH2A or a functional variant or fragment thereof.

[653] Embodiment 148. The system of any one of embodiments 132-147, wherein the target sequence comprises a nucleotide sequence within **SEQ ID NO: 266**.

[654] Embodiment 149. The system of embodiment 148, wherein the spacer sequence is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 4**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6, 13, 47**, and **228-231**.

[655] Embodiment 150. The system of embodiment 149, wherein the guide nucleic acid comprises a spacer sequence that is at least 90% identical to any one of **SEQ ID NO: 348, 558, 401, 409, 473** and **486**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6,** and **228-230**.

[656] Embodiment 151. The system of embodiment 149, wherein the guide nucleic acid comprises a spacer sequence that is at least 90% identical to any one of **SEQ ID NO: 592, 600** and **602**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 231**.

[657] Embodiment 152. The system of embodiment 149, wherein the guide nucleic acid comprises a nucleotide sequence that is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 9**.

[658] Embodiment 153. The system of embodiment 152, wherein the guide nucleic acid comprises a nucleotide sequence that is at least 90% identical to any one of **SEQ ID NO: 652, 685, 705, 713, 777** and 790, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6,** and **228-230**.

[659] Embodiment 154. The system of embodiment 152, wherein the guide nucleic acid comprises a nucleotide sequence that is at least 90% identical to any one of **SEQ ID NO: 856, 864** and **866**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 231**.

[660] Embodiment 155. The system of any one of embodiments 132-147, wherein the target sequence comprises a nucleotide sequence within **SEQ ID NO: 1795**.

[661] Embodiment 156. The system of embodiment 155, wherein the spacer sequence is at least 90% identical to any one of nucleotide sequences recited in **TABLE 5**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6, 13, 47,** and **228-231**.

[662] Embodiment 157. The system of embodiment 150 or 153, wherein the effector protein comprises one or more amino acid substitutions independently selected from E109R, H208R, K184R, K38R, L182R, Q183R, S108R, S198R, T114R, or a combination thereof.

[663] Embodiment 158. The system of embodiment 151 or 154, wherein the effector protein comprising one or more amino acid substitutions independently selected from K58W, I80K, N193K, S209F, A218K, E225K, N286K, M295W, M298L, A306K, Y315M, or a combination thereof.

[664] Embodiment 159. The system of embodiment 151 or 154, wherein the effector protein comprises one or more amino acid substitutions independently selected from I80R, T84R, K105R, G210R, C202R, A218R, D220R, E225R, C246R, Q360R, or a combination thereof.

[665] Embodiment 160. The composition of any one of embodiments 69-93, the pharmaceutical composition of embodiment 94, the system of embodiment 117 or 118, the method of any one of embodiments 119-124, or the cells of any one of embodiments 125-131, wherein the target nucleic acid is within the human albumin gene.

[666] Embodiment 161. The composition, system, method or cells of embodiment 160, wherein the target nucleic acid is at least partially within a targeted intron within the human albumin gene.

[667] Embodiment 162. The composition, system, method or cells of embodiment 160 or 161, wherein at least a portion of the target nucleic acid that the guide nucleic acid binds to comprises about 30 nucleotides to about 150 nucleotides adjacent to: the start of the intron 1, the end of the intron 1, or both.

[668] Embodiment 163. A method of inserting a donor nucleic acid within a safe harbor loci using CasPhi.12 L26R or variants thereof.

[669] Embodiment 164. The method of embodiment 163, wherein the donor nucleic acid comprises a nucleotide sequence that encodes a functional human protein into a safe harbor locus.

[670] Embodiment 165. The composition of any one of embodiments 1-31, wherein the guide nucleic acid comprises a repeat sequence of **SEQ ID NO: 252** or **1848**, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6** and **228-230**, and wherein the target sequence is adjacent to a PAM of **SEQ ID NO: 234-237**.

[671] Embodiment 166. The composition of any one of embodiments 1-31, wherein the guide nucleic acid comprises a repeat sequence of **SEQ ID NO: 253**, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 13**, and wherein the target sequence is adjacent to a PAM of **SEQ ID NO: 238, 240-247**.

[672] Embodiment 167. The composition of any one of embodiments 1-31, wherein the guide nucleic acid comprises a handle sequence of **SEQ ID NO: 259-261**, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 13**, and wherein the target sequence is adjacent to a PAM of **SEQ ID NO: 238, 240-247**.

[673] Embodiment 168. The composition of any one of embodiments 1-31, wherein the guide nucleic acid comprises a repeat sequence of **SEQ ID NO: 254**, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 47**, and wherein the target sequence is adjacent to a PAM of **SEQ ID NO: 241-244**.

[674] Embodiment 169. The composition of any one of embodiments 1-31, wherein the guide nucleic acid comprises a repeat sequence of **SEQ ID NO: 255** or **1789**, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 231**, and wherein the target sequence is adjacent to a PAM of any one of **SEQ ID NO: 245-248, 1829,** and **1834-1844**.

[675] Embodiment 170. The composition of any one of embodiments 1-31, wherein the guide nucleic acid comprises any one of handle sequences of **SEQ ID NO: 262-264**, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 231**, and wherein the target sequence is adjacent to a PAM of any one of **SEQ ID NO: 245-248, 1829,** and **1834-1844**.

SEQUENCES AND TABLES

[676] **TABLE 1A** provides exemplary amino acid sequence for effector proteins.

TABLE 1A. Exemplary Amino Acid Sequence of Effector Protein

Name	SEQ ID NO:	Effector Protein Amino Acid Sequence
Active CasPhi.12	**6**	MIKPTVSQFLTPGFKLIRNHSRTAGLKLKNEGEEACKKFVRENEIPKDECPNFQGGPAIANIIAKSREFTEWEIYQSSLAIQE VIFTLPKDKLPEPILKEEWRAQWLSEHGLDTVPYKEAAGLNLIIKNAVNTYKGVQVKVDNKNKNNLAKINRKNEIAKLN GEQEISFEEIKAFDDKGYLLQKPSPNKSIYCYQSVSPKPFITSKYHNVNLPEEYIGYYRKSNEPIVSPYQFDRLRIPIGEPGY VPKWQYTFLSKKENKRRKLSKRIKNVSPILGIICIKKDWCVFDMRGLLRTNHWKKYHKPTDSINDLFDYFTGDPVIDTKA NVVRFRYKMENGIVNYKPVREKKGKELLENICDQNGSCKLATVDVGQNNPVAIGLFELKKVNGELTKTLISRHPTPIDFC NKITAYRERYDKLESSIKLDAIKQLTSEQKIEVDNYNNNFTPQNTKQIVCSKLNINPNDLPWDKMISGTHFISEKAQVSNK SEIYFTSTDKGKTKDVMKSDYKWFQDYKPKLSKEVRDALSDIEWRLRRESLEFNKLSKSREQDARQLANWISSMCDVIG IENLVKKNNFFGGSGKREPGWDNFYKPKKENRWWINAIHKALTELSQNKGKRVILLPAMRTSITCPKCKYCDSKNRNGE KFNCLKCGIELNADIDVATENLATVAITAQSMPKPTCERSGDAKKPVRARKAKAPEFHDKLAPSYTVVLREAV
Dead CasPhi.12 (D369A)	**7**	MIKPTVSQFLTPGFKLIRNHSRTAGLKLKNEGEEACKKFVRENEIPKDECPNFQGGPAIANIIAKSREFTEWEIYQSSLAIQE VIFTLPKDKLPEPILKEEWRAQWLSEHGLDTVPYKEAAGLNLIIKNAVNTYKGVQVKVDNKNKNNLAKINRKNEIAKLN GEQEISFEEIKAFDDKGYLLQKPSPNKSIYCYQSVSPKPFITSKYHNVNLPEEYIGYYRKSNEPIVSPYQFDRLRIPIGEPGY VPKWQYTFLSKKENKRRKLSKRIKNVSPILGIICIKKDWCVFDMRGLLRTNHWKKYHKPTDSINDLFDYFTGDPVIDTKA NVVRFRYKMENGIVNYKPVREKKGKELLENICDQNGSCKLATVAVGQNNPVAIGLFELKKVNGELTKTLISRHPTPIDFC NKITAYRERYDKLESSIKLDAIKQLTSEQKIEVDNYNNNFTPQNTKQIVCSKLNINPNDLPWDKMISGTHFISEKAQVSNK SEIYFTSTDKGKTKDVMKSDYKWFQDYKPKLSKEVRDALSDIEWRLRRESLEFNKLSKSREQDARQLANWISSMCDVIG IENLVKKNNFFGGSGKREPGWDNFYKPKKENRWWINAIHKALTELSQNKGKRVILLPAMRTSITCPKCKYCDSKNRNGE KFNCLKCGIELNADIDVATENLATVAITAQSMPKPTCERSGDAKKPVRARKAKAPEFHDKLAPSYTVVLREAV
Dead CasPhi.12 (D369N)	**8**	MIKPTVSQFLTPGFKLIRNHSRTAGLKLKNEGEEACKKFVRENEIPKDECPNFQGGPAIANIIAKSREFTEWEIYQSSLAIQE VIFTLPKDKLPEPILKEEWRAQWLSEHGLDTVPYKEAAGLNLIIKNAVNTYKGVQVKVDNKNKNNLAKINRKNEIAKLN GEQEISFEEIKAFDDKGYLLQKPSPNKSIYCYQSVSPKPFITSKYHNVNLPEEYIGYYRKSNEPIVSPYQFDRLRIPIGEPGY VPKWQYTFLSKKENKRRKLSKRIKNVSPILGIICIKKDWCVFDMRGLLRTNHWKKYHKPTDSINDLFDYFTGDPVIDTKA NVVRFRYKMENGIVNYKPVREKKGKELLENICDQNGSCKLATVNVGQNNPVAIGLFELKKVNGELTKTLISRHPTPIDFC NKITAYRERYDKLESSIKLDAIKQLTSEQKIEVDNYNNNFTPQNTKQIVCSKLNINPNDLPWDKMISGTHFISEKAQVSNK SEIYFTSTDKGKTKDVMKSDYKWFQDYKPKLSKEVRDALSDIEWRLRRESLEFNKLSKSREQDARQLANWISSMCDVIG IENLVKKNNFFGGSGKREPGWDNFYKPKKENRWWINAIHKALTELSQNKGKRVILLPAMRTSITCPKCKYCDSKNRNGE KFNCLKCGIELNADIDVATENLATVAITAQSMPKPTCERSGDAKKPVRARKAKAPEFHDKLAPSYTVVLREAV

Dead CasPhi.12 (E567A)	**9**	MIKPTVSQFLTPGFKLIRNHSRTAGLKLKNEGEEACKKFVRENEIPKDECPNFQGGPAIANIIAKSREFTEWEIYQSSLAIQE VIFTLPKDKLPEPILKEEWRAQWLSEHGLDTVPYKEAAGLNLIIKNAVNTYKGVQVKVDNKNKNNLAKINRKNEIAKLN GEQEISFEEIKAFDDKGYLLQKPSPNKSIYCYQSVSPKPFITSKYHNVNLPEEYIGYYRKSNEPIVSPYQFDRLRIPIGEPGY VPKWQYTFLSKKENKRRKLSKRIKNVSPILGIICIKKDWCVFDMRGLLRTNHWKKYHKPTDSINDLFDYFTGDPVIDTKA NVVRFRYKMENGIVNYKPVREKKGKELLENICDQNGSCKLATVDVGQNNPVAIGLFELKKVNGELTKTLISRHPTPIDFC NKITAYRERYDKLESSIKLDAIKQLTSEQKIEVDNYNNNFTPQNTKQIVCSKLNINPNDLPWDKMISGTHFISEKAQVSNK SEIYFTSTDKGKTKDVMKSDYKWFQDYKPKLSKEVRDALSDIEWRLRRESLEFNKLSKSREQDARQLANWISSMCDVIG IANLVKKNNFFGGSGKREPGWDNFYKPKKENRWWINAIHKALTELSQNKGKRVILLPAMRTSITCPKCKYCDSKNRNGE KFNCLKCGIELNADIDVATENLATVAITAQSMPKPTCERSGDAKKPVRARKAKAPEFHDKLAPSYTVVLREAV
Dead CasPhi.12 (E567Q)	**10**	MIKPTVSQFLTPGFKLIRNHSRTAGLKLKNEGEEACKKFVRENEIPKDECPNFQGGPAIANIIAKSREFTEWEIYQSSLAIQE VIFTLPKDKLPEPILKEEWRAQWLSEHGLDTVPYKEAAGLNLIIKNAVNTYKGVQVKVDNKNKNNLAKINRKNEIAKLN GEQEISFEEIKAFDDKGYLLQKPSPNKSIYCYQSVSPKPFITSKYHNVNLPEEYIGYYRKSNEPIVSPYQFDRLRIPIGEPGY VPKWQYTFLSKKENKRRKLSKRIKNVSPILGIICIKKDWCVFDMRGLLRTNHWKKYHKPTDSINDLFDYFTGDPVIDTKA NVVRFRYKMENGIVNYKPVREKKGKELLENICDQNGSCKLATVDVGQNNPVAIGLFELKKVNGELTKTLISRHPTPIDFC NKITAYRERYDKLESSIKLDAIKQLTSEQKIEVDNYNNNFTPQNTKQIVCSKLNINPNDLPWDKMISGTHFISEKAQVSNK SEIYFTSTDKGKTKDVMKSDYKWFQDYKPKLSKEVRDALSDIEWRLRRESLEFNKLSKSREQDARQLANWISSMCDVIG IQNLVKKNNFFGGSGKREPGWDNFYKPKKENRWWINAIHKALTELSQNKGKRVILLPAMRTSITCPKCKYCDSKNRNGE KFNCLKCGIELNADIDVATENLATVAITAQSMPKPTCERSGDAKKPVRARKAKAPEFHDKLAPSYTVVLREAV
Dead CasPhi.12 (D658A)	**11**	MIKPTVSQFLTPGFKLIRNHSRTAGLKLKNEGEEACKKFVRENEIPKDECPNFQGGPAIANIIAKSREFTEWEIYQSSLAIQE VIFTLPKDKLPEPILKEEWRAQWLSEHGLDTVPYKEAAGLNLIIKNAVNTYKGVQVKVDNKNKNNLAKINRKNEIAKLN GEQEISFEEIKAFDDKGYLLQKPSPNKSIYCYQSVSPKPFITSKYHNVNLPEEYIGYYRKSNEPIVSPYQFDRLRIPIGEPGY VPKWQYTFLSKKENKRRKLSKRIKNVSPILGIICIKKDWCVFDMRGLLRTNHWKKYHKPTDSINDLFDYFTGDPVIDTKA NVVRFRYKMENGIVNYKPVREKKGKELLENICDQNGSCKLATVDVGQNNPVAIGLFELKKVNGELTKTLISRHPTPIDFC NKITAYRERYDKLESSIKLDAIKQLTSEQKIEVDNYNNNFTPQNTKQIVCSKLNINPNDLPWDKMISGTHFISEKAQVSNK SEIYFTSTDKGKTKDVMKSDYKWFQDYKPKLSKEVRDALSDIEWRLRRESLEFNKLSKSREQDARQLANWISSMCDVIG IENLVKKNNFFGGSGKREPGWDNFYKPKKENRWWINAIHKALTELSQNKGKRVILLPAMRTSITCPKCKYCDSKNRNGE KFNCLKCGIELNAADIDVATENLATVAITAQSMPKPTCERSGDAKKPVRARKAKAPEFHDKLAPSYTVVLREAV
Dead CasPhi.12 (D658N)	**12**	MIKPTVSQFLTPGFKLIRNHSRTAGLKLKNEGEEACKKFVRENEIPKDECPNFQGGPAIANIIAKSREFTEWEIYQSSLAIQE VIFTLPKDKLPEPILKEEWRAQWLSEHGLDTVPYKEAAGLNLIIKNAVNTYKGVQVKVDNKNKNNLAKINRKNEIAKLN GEQEISFEEIKAFDDKGYLLQKPSPNKSIYCYQSVSPKPFITSKYHNVNLPEEYIGYYRKSNEPIVSPYQFDRLRIPIGEPGY VPKWQYTFLSKKENKRRKLSKRIKNVSPILGIICIKKDWCVFDMRGLLRTNHWKKYHKPTDSINDLFDYFTGDPVIDTKA NVVRFRYKMENGIVNYKPVREKKGKELLENICDQNGSCKLATVDVGQNNPVAIGLFELKKVNGELTKTLISRHPTPIDFC NKITAYRERYDKLESSIKLDAIKQLTSEQKIEVDNYNNNFTPQNTKQIVCSKLNINPNDLPWDKMISGTHFISEKAQVSNK SEIYFTSTDKGKTKDVMKSDYKWFQDYKPKLSKEVRDALSDIEWRLRRESLEFNKLSKSREQDARQLANWISSMCDVIG

		IENLVKKNNFFGGSGKREPGWDNFYKPKKENRWWINAIHKALTELSQNKGKRVILLPAMRTSITCPKCKYCDSKNRNGE KFNCLKCGIELNANIDVATENLATVAITAQSMPKPTCERSGDAKKPVRARKAKAPEFHDKLAPSYTVVLREAV
Cas14a.1	**13**	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACSKHLKVAAYCTTQVERNACLFCKARKLD DKFYQKLRGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSDVCYTRAAELFKNAAIA SGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFE QVQKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKIGEKSAWMLNLSIDVPKIDKGVDPSII GGIDVGVKSPLVCAINNAFSRYSISDNDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERW ACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRKVAPNNTSKTCSKCGHLN NYFNFEYRKKNKFPHFKCEKCNFKENADYNAALNISNPKLKSTKEEP
Cas14a.2 80852	**14**	MAGKKKDKDVINKTLSVRIIRPRYSDDIEKEISDEKAKRKQDGKTGELDRAFFSELKSRNPDIITNDELFPLFTEIQKNLTEI YNKSISLLYMKLIVEEEGGSTASALSAGPYKECKARFNSYISLGLRQKIQSNFRRKELKGFQVSLPTAKSDRFPIPFCHQVE NGKGGFKVYETGDDFIFEVPLIKYTATNKKSTSGKNYTKVQLNNPPVPMNVPLLLSTMRRRQTKKGMQWNKDEGTNA ELRRVMSGEYKVSYAEIIRRTRFGKHDDWFVNFSIKFKNKTDELNQNVRGGIDIGVSNPLVCAVTNGLDRYIVANNDIM AFNERAMARRRTLLRKNRFKRSGHGAKNKLEPITVLTEKNERFRKSILQRWAREVAEFFKRTSASVVNMEDLSGITERE DFFSTKLRTTWNYRLMQTTIENKLKEYGIAVNYISPKYTSQTCHSCGKRNDYFTFSYRSENNYPPFECKECNKVKCNADF NAAKNIALKVVL
CasPhi.32	**47**	VPDKKETPLVALCKKSFPGLRFKKHDSRQAGRILKSKGEGAAVAFLEGKGGTTQPNFKPPVKCNIVAMSRPLEEWPIYK ASVVIQKYVYAQSYEEFKATDPGKSEAGLRAWLKATRVDTDGYFNVQGLNLIFQNARATYEGVLKKVENRNSKKVAKI EQRNEHRAERGLPLLTLDEPETALDETGHLRHRPGINCSVFGYGYQHMKLKPYVPGSIPGVTGYSRDPSTPIAACGVDRLEIP EGQPGYVPPWDRENLSVKKHRRKRASWARSRGGAIDDNMLLAVVRVADDWALLDLRGLLRNTQYRKLLDRSVPVTIE SLLNLVTNDPTLSVVKKPGKPVRYTATLIYKQGVVPVVKAKVVKGSYVSKMLDDTTETFSLVGVDLGVNNLIAANALRI RPGKCVERLQAFTLPEQTVEDFFRFRKAYDKHQENLRLAAVRSLTAEQQAEVLALDTFGPEQAKMQVCGHLGLSVDEV PWDKVNSRSSILSDLAKERGVDDTLYMFPFFKGKGKKRKTEIRKRWDVNWAQHFRPQLTSETRKALNEAKWEAERNSS KYHQLSIRKKELSRHCVNYVIRTAEKRAQCGKVIVAVEDLHHSFRRGGKGSRKSGWGGFFAAKQEGRWLMDALFGAF CDLAVHRGYRVIKVDPYNTSRTCPECGHCDKANRDRVVNREAFICVCCGYRGNADIDVAAYNIAMVAITGVSLRKAARA SVASTPLESLAAE
CasPhi.12 L26R	**228**	MIKPTVSQFLTPGFKLIRNHSRTAGRKLKNEGEEACKKFVRENEIPKDECPNFQGGPAIANIIAKSREFTEWEIYQSSLAIQE VIFTLPKDKLPEPILKEEWRAQWLSEHGLDTVPYKEAAGLNLIIKNAVNTYKGVQVKVDNKNKNNLAKINRKNEIAKLN GEQEISFEEIKAFDDKGYLLQKPSPNKSIYCYQSVSPKPFITSKYHNVNLPEEYIGYYRKSNEPIVSPYQFDRLRIPIGEPGY VPKWQYTFLSKKENKRRKLSKRIKNVSPILGIICIKKDWCVFDMRGLLRTNHWKKYHKPTDSINDLFDYFTGDPVIDTKA NVVRFRYKMENGIVNYKPVREKKGKELLENICDQNGSCKLATVDVGQNNPVAIGLFELKKVNGELTKTLISRHPTPIDFC NKITAYRERYDKLESSIKLDAIKQLTSEQKIEVDNYNNNFTPQNTKQIVCSKLNINPNDLPWDKMISGTHFISEKAQVSNK SEIYFTSTDKGKTKDVMKSDYKWFQDYKPKLSKEVRDALSDIEWRLRRESLEFNKLSKSREQDARQLANWISSMCDVIG IENLVKKNNFFGGSGKREPGWDNFYKPKKENRWWINAIHKALTELSQNKGKRVILLPAMRTSITCPKCKYCDSKNRNGE KFNCLKCGIELNADIDVATENLATVAITAQSMPKPTCERSGDAKKPVRARKAKAPEFHDKLAPSYTVVLREAV

CasPhi.12 L26H	**229**	MIKPTVSQFLTPGFKLIRNHSRTAGHKLKNEGEEACKKFVRENEIPKDECPNFQGGPAIANIIAKSREFTEWEIYQSSLAIQ EVIFTLPKDKLPEPILKEEWRAQWLSEHGLDTVPYKEAAGLNLIIKNAVNTYKGVQVKVDNKNKNNLAKINRKNEIAKL NGEQEISFEEIKAFDDKGYLLQKPSPNKSIYCYQSVSPKPFITSKYHNVNLPEEYIGYYRKSNEPIVSPYQFDRLRIPIGEPG YVPKWQYTFLSKKENKRRKLSKRIKNVSPILGIICIKKDWCVFDMRGLLRTNHWKKYHKPTDSINDLFDYFTGDPVIDTK ANVVRFRYKMENGIVNYKPVREKKGKELLENICDQNGSCKLATVDVGQNNPVAIGLFELKKVNGELTKTLISRHPTPIDF CNKITAYRERYDKLESSIKLDAIKQLTSEQKIEVDNYNNNFTPQNTKQIVCSKLNINPNDLPWDKMISGTHFISEKAQVSN KSEIYFTSTDKGKTKDVMKSDYKWFQDYKPKLSKEVRDALSDIEWRLRRESLEFNKLSKSREQDARQLANWISSMCDVI GIENLVKKNNFFGGSGKREPGWDNFYKPKKENRWWINAIHKALTELSQNKGKRVILLPAMRTSITCPKCKYCDSKNRNG EKFNCLKCGIELNADIDVATENLATVAITAQSMPKPTCERSGDAKKPVRARKAKAPEFHDKLAPSYTVVLREAV
CasPhi.12 L26K	**230**	MIKPTVSQFLTPGFKLIRNHSRTAGKKLKNEGEEACKKFVRENEIPKDECPNFQGGPAIANIIAKSREFTEWEIYQSSLAIQ EVIFTLPKDKLPEPILKEEWRAQWLSEHGLDTVPYKEAAGLNLIIKNAVNTYKGVQVKVDNKNKNNLAKINRKNEIAKL NGEQEISFEEIKAFDDKGYLLQKPSPNKSIYCYQSVSPKPFITSKYHNVNLPEEYIGYYRKSNEPIVSPYQFDRLRIPIGEPG YVPKWQYTFLSKKENKRRKLSKRIKNVSPILGIICIKKDWCVFDMRGLLRTNHWKKYHKPTDSINDLFDYFTGDPVIDTK ANVVRFRYKMENGIVNYKPVREKKGKELLENICDQNGSCKLATVDVGQNNPVAIGLFELKKVNGELTKTLISRHPTPIDF CNKITAYRERYDKLESSIKLDAIKQLTSEQKIEVDNYNNNFTPQNTKQIVCSKLNINPNDLPWDKMISGTHFISEKAQVSN KSEIYFTSTDKGKTKDVMKSDYKWFQDYKPKLSKEVRDALSDIEWRLRRESLEFNKLSKSREQDARQLANWISSMCDVI GIENLVKKNNFFGGSGKREPGWDNFYKPKKENRWWINAIHKALTELSQNKGKRVILLPAMRTSITCPKCKYCDSKNRNG EKFNCLKCGIELNADIDVATENLATVAITAQSMPKPTCERSGDAKKPVRARKAKAPEFHDKLAPSYTVVLREAV
CasM.265466	**231**	MSVLTRKVQLIPVGDKEERDRVYKYLRDGIEAQNRAMNLYMSGLYFAAINEASKEDRKELNQLYSRIATSSKGSAYTTD IEFPTGLASTSTLSMAVRQDFTKSLKDGLMYGRVSLPTYRKDNPLFVDVRFVALRGTKQKYNGLYHEYKSHTEFLDNLY SSDLKVYIKFANDITFQVIFGNPRKSSALRSEFQNIFEEYYKVCQSSIQFSGTKIILNMAMDIPDKEIELDEDVCVGVDLGIA IPAVCALNKNRYSRVSIGSKEDFLRVRTKIRNQRKRLQTNLKSSNGGHGRKKKMKPMDRFRDYEANWVQNYNHYVSR QVVDFAVKNKAKYINLENLEGIRDDVKNEWLLSNWSYYQLQQYITYKAKTYGIEVRKINPYHTSQRCSCCGYEDAGNR PKKEKGQAYFKCLKCGEEMNADFNAARNIAMSTEFQSGKKTKKQKKEQHENK

[677] **TABLE 1B** provides exemplary effector proteins.

TABLE 1B. Exemplary Effector Proteins

Name	SEQ ID NO:
Active SpyCas9	**1**
Dead SpyCas9	**2**
Dead PspCas13b	**3**
Nickase SpyCas9 (D10A)	**4**
Nickase SpyCas9 (H840A)	**5**
Nickase SpyCas9	**15**
Cas13	**16**
CasPhi.1	**17**
CasPhi.2	**18**
CasPhi.3	**19**
CasPhi.4	**20**
CasPhi.5	**21**
CasPhi.6	**22**
CasPhi.7	**23**
CasPhi.8	**24**
CasPhi.9	**25**
CasPhi.10	**26**
CasPhi.11	**27**
CasPhi.13	**28**
CasPhi.14	**29**
CasPhi.15	**30**
CasPhi.16	**31**
CasPhi.17	**32**
CasPhi.18	**33**
CasPhi.19	**34**
CasPhi.20	**35**
CasPhi.21	**36**
CasPhi.22	**37**
CasPhi.23	**38**
CasPhi.24	**39**
CasPhi.25	**40**

Name	SEQ ID NO:
CasPhi.26	**41**
CasPhi.27	**42**
CasPhi.28	**43**
CasPhi.29	**44**
CasPhi.30	**45**
CasPhi.31	**46**
CasPhi.33	**48**
CasPhi.34	**49**
CasPhi.35	**50**
CasPhi.36	**51**
CasPhi.37	**52**
CasPhi.38	**53**
CasPhi.39	**54**
CasPhi.41	**55**
CasPhi.42	**56**
CasPhi.43	**57**
CasPhi.44	**58**
CasPhi.45	**59**
CasPhi.46	**60**
CasPhi.47	**61**
CasPhi.48	**62**
CasPhi.49	**63**
Cas14 ortholog 1	**64**
Cas14 ortholog 3	**65**
Cas14 ortholog 4	**66**
Cas14 ortholog 5	**67**
Cas14 ortholog 6	**68**
Cas14 ortholog 7	**69**
Cas14 ortholog 9	**70**
Cas14 ortholog 10	**71**
Cas14 ortholog 11	**72**
Cas14 ortholog 12	**73**
Cas14 ortholog 13	**74**

Name	SEQ ID NO:
Cas14 ortholog 14	**75**
Cas14 ortholog 15	**76**
Cas14 ortholog 16	**77**
Cas14 ortholog 17	**78**
Cas14 ortholog 18	**79**
Cas14 ortholog 19	**80**
Cas14 ortholog 20	**81**
Cas14 ortholog 21	**82**
Cas14 ortholog 22	**83**
Cas14 ortholog 23	**84**
Cas14 ortholog 24	**85**
Cas14 ortholog 25	**86**
Cas14 ortholog 26	**87**
Cas14 ortholog 27	**88**
Cas14 ortholog 28	**89**
Cas14 ortholog 29	**90**
Cas14 ortholog 30	**91**
Cas14 ortholog 31	**92**
Cas14 ortholog 32	**93**
Cas14 ortholog 33	**94**
Cas14 ortholog 34	**95**
Cas14 ortholog 35	**96**
Cas14 ortholog 36	**97**
Cas14 ortholog 37	**98**
Cas14 ortholog 38	**99**
Cas14 ortholog 39	**100**
Cas14 ortholog 40	**101**
Cas14 ortholog 41	**102**
Cas14 ortholog 42	**103**
Cas14 ortholog 43	**104**
Cas14 ortholog 44	**105**
Cas14 ortholog 45	**106**
Cas14 ortholog 46	**107**

Name	SEQ ID NO:
Cas14 ortholog 47	**108**
Cas14 ortholog 48	**109**
Cas14 ortholog 49	**110**
Cas14 ortholog 50	**111**
Cas14 ortholog 51	**112**
Cas14 ortholog 52	**113**
Cas14 ortholog 53	**114**
Cas14 ortholog 54	**115**
Cas14 ortholog 55	**116**
Cas14 ortholog 56	**117**
Cas14 ortholog 57	**118**
Cas14 ortholog 58	**119**
Cas14 ortholog 59	**120**
Cas14 ortholog 60	**121**
Cas14 ortholog 61	**122**
Cas14 ortholog 62	**123**
Cas14 ortholog 63	**124**
Cas14 ortholog 64	**125**
Cas14 ortholog 65	**126**
Cas14 ortholog 66	**127**
Cas14 ortholog 67	**128**
Cas14 ortholog 68	**129**
Cas14 ortholog 69	**130**
Cas14 ortholog 70	**131**
Cas14 ortholog 71	**132**
Cas14 ortholog 72	**133**
Cas14 ortholog 73	**134**
Cas14 ortholog 74	**135**
Cas14 ortholog 75	**136**
Cas14 ortholog 76	**137**
Cas14 ortholog 77	**138**
Cas14 ortholog 78	**139**
Cas14 ortholog 79	**140**

Name	SEQ ID NO:
Cas14 ortholog 80	**141**
Cas14 ortholog 81	**142**
Cas14 ortholog 82	**143**
Cas14 ortholog 83	**144**
Cas14 ortholog 84	**145**
Cas14 ortholog 85	**146**
Cas14 ortholog 86	**147**
Cas14 ortholog 87	**148**
Cas14 ortholog 88	**149**
Cas14 ortholog 89	**150**
Cas14 ortholog 90	**151**
Cas14 ortholog 91	**152**
Cas14 ortholog 92	**153**
Cas14 ortholog 93	**154**
Cas14 ortholog 94	**155**
Cas14 ortholog 95	**156**
Cas14 ortholog 96	**157**
Cas14 ortholog 97	**158**
CasM.298706	**159**
CasM.280604	**160**
CasM.281060	**161**
CasM.284933	**162**
CasM.287908	**163**
CasM.288518	**164**
CasM.293891	**165**
CasM.294270	**166**
CasM.294491	**167**
CasM.295047	**168**
CasM.299588	**169**
CasM.277328	**170**
CasM.297894	**171**
CasM.291449	**172**
CasM.297599	**173**

Name	SEQ ID NO:
CasM.286588	**174**
CasM.286910	**175**
CasM.292335	**176**
CasM.293576	**177**
CasM.294537	**178**
CasM.298538	**179**
CasM.19924	**180**
CasM.19952	**181**
CasM.19952 (D267A)	**182**
CasM.19952 (D267N)	**183**
CasM.19952 (E363Q)	**184**
CasM.124070 (D326A)	**185**
CasM.274559	**186**
CasM.286251	**187**
CasM.286251 (D267A)	**188**
CasM.288480	**189**
CasM.288668	**190**
CasM.289206	**191**
CasM.290598	**192**
CasM.290816	**193**
CasM.295071	**194**
CasM.295231	**195**
CasM.292139	**196**
CasM.279423	**197**
CasM.20054	**198**
CasM.282673	**199**
CasM.282952	**200**
CasM.283262	**201**
CasM.284833	**202**
CasM.287700	**203**
CasM.291507	**204**
CasM.293410	**205**
CasM.295105	**206**

Name	SEQ ID NO:
CasM.295187	**207**
CasM.295929	**208**
Cas13 ortholog 1	**209**
Cas13 ortholog 2	**210**
Cas13 ortholog 3	**211**
Cas13 ortholog 4	**212**
Cas13 ortholog 5	**213**
CasM.1584	**214**
CasM.1730	**215**
CasM.1770	**216**
CasM.1816	**217**
CasM.1862939	**218**
CasM.1862895	**219**
CasM.1862903	**220**
CasM.1862909	**221**
CasM.1862917	**222**
CasM.1862921	**223**
CasM.1862947	**224**
CasM.1422	**225**
CasM.1740	**226**

[678] **TABLE 2** provides exemplary nuclear localization sequences.

TABLE 2. Exemplary Nuclear Localization Sequences

SEQ ID NO:	Description	SEQUENCES
232	NLS	PKKKRKVGIHGVPAA
233	NLS	KRPAATKKAGQAKKKK
1809	NLS	KR(K/R)R
1810	NLS	(P/R)XXKR(^D/E)(K/R)
1811	NLS	KRX(W/F/Y)XXAF
1812	NLS	(R/P)XXKR(K/R)(^D/E)
1813	NLS	LGKR(K/R)(W/F/Y)
1814	NLS	$KRX_{10}K(K/R)(K/R)$
1815	NLS	K(K/R)RK
1816	NLS	$KRX_{11}K(K/R)(K/R)$
1817	NLS	$KRX_{12}K(K/R)(K/R)$
1818	NLS	$KRX_{10}K(K/R)X(K/R)$
1819	NLS	$KRX_{11}K(K/R)X(K/R)$
1820	NLS	$KRX_{12}K(K/R)X(K/R)$
1821	NLS	APKKKRKVGIHGVPAA
1822	EEP	GLFXALLXLLXSLWXLLLXA
1823	EEP	GLFHALLHLLHSLWHLLLHA

wherein X is any naturally occurring amino acid; and ^D/E is any naturally occurring amino acid except Asp or Glu

[679] **TABLE 3** provides exemplary PAM sequences.

TABLE 3: PAM Sequences

Effector Protein SEQ ID NO:	SEQ ID NO:	PAM Sequence (5' → 3')
6-12, 228-230	234	TTG
6-12, 228-230	235	TTA
6-12, 228-230	236	TTT
6-12, 228-230	237	TTC
13	238	TTTA

Effector Protein SEQ ID NO:	SEQ ID NO:	PAM Sequence (5' → 3')
13	240	NTTN
13, 47	241	GTTG
13, 47	242	GTTC
13, 47	243	GTTT
13, 47	244	GTTA
13, 231	245	TCTG
13, 231	246	TATG
13, 231	247	TTTG
231	248	TGTG
181-184	249	NTCG
187-188	250	RTTR
185	251	TTTR
181-184	1791	ATCG
181-184	1792	CTCG
181-184	1793	GTCG
181-184	1794	TTCG
231	1829	TNTR
231	1834	NNTN
231	1835	ANTR
231	1836	CNTR
231	1837	GNTR
231	1838	TNAR
231	1839	TNCR
231	1840	TNGR
231	1841	TNTC
231	1842	TNTT
231	1843	VNTY
231	1844	TNVY
*wherein each N is any nucleotide, each R is A or G, and each V is A, C or G		

[680] **TABLE 4** provides exemplary spacer sequences that targets albumin safe harbor.

TABLE 4: Exemplary Spacer Sequences Targeting Albumin Safe Harbor

Seq ID No:	Spacer sequence (5' → 3'), shown as RNA	Seq ID No:	Spacer sequence (5' → 3'), shown as RNA	Seq ID No:	Spacer sequence (5' → 3'), shown as RNA	Seq ID No:	Spacer sequence (5' → 3'), shown as RNA
301	AACAAUAGAAAAAUGGA	377	CACUUUCCUUAGUGCGC	453	UUAAAAUAGUAUUCUUG	529	GGAUAGUUAUGAAUUCAAUC
302	CUGGGAAAAUAGAAUAA	378	AGUGACUGUAAUUUUCU	454	UAAAAUAGUAUUCUUGG	530	AAUAAAGCAUAGUGCAAUGG
303	ACUGGGAAAAUAGAAUA	379	GCACUUUCCUUAGUGCG	455	AACAGAAGAAUAAUUCA	531	UGGUUUUUAAAUAAAGCAUA
304	UACUGGGAAAAUAGAAU	380	GUGACUGUAAUUUUCUU	456	AAAAUAGUAUUCUUGGU	532	UGAGAUCAACAGCACAGGUU
305	UUCUAUUGUUCAACUUU	381	CUUUGCACUUUCCUUAG	457	AAACAGAAGAAUAAUUC	533	UUUAAAAACCACAAAACCUG
306	UUUUACUGGGAAAAUAG	382	UCUUUUGCGCACUAAGG	458	AAAUAGUAUUCUUGGUA	534	AAAACCACAAAACCUGUGCU
307	UCUAUUGUUCAACUUUU	383	CUUUUGCGCACUAAGGA	459	UGCCUUUAAACAGAAGA	535	UAUUUAUUUUCAUUUUAGUC
308	AUUUUACUGGGAAAAUA	384	UUUUGCGCACUAAGGAA	460	UUCUGCCUUUAAACAGA	536	UUUUCAUUUUAGUCUGUCUU
309	CUAUUGUUCAACUUUUA	385	UGCGCACUAAGGAAAGU	461	CUUCUGCCUUUAAACAG	537	GUGUCUAUCAACAGCAACCA
310	UAUUGUUCAACUUUUAU	386	AGUCACUCUAAGUUACU	462	UUGGUAAUUGAAUUAUU	538	GUCUGUCUUCUUGGUUGCUG
311	UUCAACUUUUAUUCUAU	387	GCGCACUAAGGAAAGUG	463	UUUCUUCUGCCUUUAAA	539	UAGCCUUAUAUUCAAACUUA
312	AACUUUUAUUCUAUUUU	388	CAGUCACUCUAAGUUAC	464	GUAAUUGAAUUAUUCUU	540	AAAAUUAUUAAAUAUUUAUA
313	CUAAAACUUUAUUUUAC	389	CGCACUAAGGAAAGUGC	465	AAUUAUUUCUUCUGCCU	541	AAUAUAAGGCUAUAAAUAUU
314	ACUAAAACUUUAUUUUA	390	UGUGAAGUUUCAGUCAC	466	AAUUAUUCUUCUGUUUA	542	AUAAUUUUUAAAAUAGUAUU
315	UAUUCUAUUUUCCCAGU	391	AACCCUAUUCUGUGAAG	467	UUCUUCUGUUUAAAGGC	543	AACAGAAGAAUAAUUCAAUU

Seq ID No:	Spacer sequence (5' → 3'), shown as RNA
316	AUUCUAUUUUCCCAGUA
317	UUCUAUUUUCCCAGUAA
318	UAUUUUCCCAGUAAAAU
319	AAGAUGCAGAGUUUAC U
320	UCCCAGUAAAAUAAAG U
321	AAAGAUGCAGAGUUU AC
322	CCCAGUAAAAUAAAGU U
323	CCAGUAAAAUAAAGUU U
324	UUUAAAGAUGCAGAG UU
325	UAGUAAACUCUGCAUC U
326	AGUAAACUCUGCAUCU U
327	GUAAACUCUGCAUCUU U
328	GAAAUAAAUGCCAAAA U
329	AGAAAUAAAUGCCAAA A
330	UAGAAAUAAAUGCCAA A

Seq ID No:	Spacer sequence (5' → 3'), shown as RNA
392	AAUCUUCAACCCUAUU C
393	GAGUGACUGAAACUUC A
394	UGAAUUCAAUCUUCAA C
395	GGAUAGUUAUGAAUU CA
396	GGGAUAGUUAUGAAU UC
397	ACAGAAUAGGGUUGA AG
398	AAGAUUGAAUUCAUA AC
399	AAUUCAUAACUAUCCC A
400	AUAACUAUCCCAAAGA C
401	AAUAAAGCAUAGUGC AA
402	AAAUAAAGCAUAGUG CA
403	UAAAUAAAGCAUAGU GC
404	UUAAAUAAAGCAUAG UG
405	UGGUUUUUAAAUAAA GC
406	GUGGUUUUUAAAUAA AG

Seq ID No:	Spacer sequence (5' → 3'), shown as RNA
468	UUCUGUUUAAAGGCAGA
469	UGUUUAAAGGCAGAAGA
470	AAAGGCAGAAGAAAUAA
471	AAGGCAGAAGAAAUAAU
472	CUACAGAAAAACUCAGG
473	GGCUCUGAUUCCUACAG
474	AACAUCAUCCUGAGUUU
475	AAAAUAUUGGGCUCUGA
476	CAAAAUAUUGGGCUCUG
477	UUUCAAAAUAUUGGGCU
478	GUUUCAAAAUAUUGGGC
479	UUCUGUAGGAAUCAGAG
480	UCUGUAGGAAUCAGAGC
481	CUGUAGGAAUCAGAGCC
482	UGCAUUUGUUUCAAAAU

Seq ID No:	Spacer sequence (5' → 3'), shown as RNA
544	AAAUAGUAUUCUUGGUAAU U
545	AAGGCAGAAGAAAUAAUUG A
546	UUUCAAAAUAUUGGGCUCU G
547	ACUUAGAUUAUGCAUUUGU U
548	UAUUUCUUUCCAUUUGACU U
549	AAACAAAUGCAUAAUCUAA G
550	ACAAUCCUUUUUUUUCUUCC
551	AACAAUAGAAAAAUGGAUU U
552	AACUUUUAUUCUAUUUUCC C
553	ACUAAAACUUUAUUUUACU G
554	UAGUAAACUCUGCAUCUUU A
555	GAAUUUUAUUAAUAAGAUA A
556	UCUUAUUAAUAAAAUUCAA A
557	AGUGACUGUAAUUUUCUUU U
558	CUUUGCACUUUCCUUAGUGC

Seq ID No:	Spacer sequence (5' → 3'), shown as RNA
331	AAAGAAUUAUUUUGGCA
332	AAGAAUUAUUUUGGC AU
333	UUUUGGCAUUUAUUUCU
334	UGGCAUUUAUUUCUAAA
335	GGCAUUUAUUUCUAAAA
336	GCAUUUAUUUCUAAAAU
337	ACAAAUACAAAAUACUA
338	AUUUCUAAAAUGGCAUA
339	UUUCUAAAAUGGCAUAG
340	UAAGACUUCACAAAUAC
341	CUAAAAUGGCAUAGUAU
342	UAAAAUGGCAUAGUAUU
343	AUAAGAUAACCUUGUAA
344	UUAAUAAGAUAACCUUG
345	AUUAAUAAGAUAACCUU
407	UGUGGUUUUUAAAUAAA
408	CACUAUGCUUUAUUUAA
409	UGAGAUCAACAGCACAG
410	AUGAGAUCAACAGCACA
411	AUUUAAAAACCACACAAA
412	UUUAAAAACCACACAAAC
413	UAUUUAUGAGAUCAACA
414	AAAAACCACACAAAACCG
415	AAAACCACACAAAACCUG
416	AUCUCAUAAAUAGAACU
417	UAUUUAUAUUUAUUUUC
418	AUAUUUAUUUUUCAUUU
419	UAUUUAUUUUUCAUUUA
420	AUUUUCAUUUUAGUCUG
421	UUUUCAUUUUAGUCUGU
483	UGUAGGAAUCAGAGCCC
484	GAUUAUGCAUUUGUUUC
485	ACUUAGAUUAUGCAUUU
486	GACUUAGAUUAUGCAUU
487	CAUUUGACUUAGAUUAU
488	CCAUUUGACUUAGAUUA
489	UUUCCAUUUGACUUAGA
490	CUUUCCAUUUGACUUAG
491	UGAAACAAAUGCAUAAU
492	GAAACAAAUGCAUAAUC
493	UAUUUCUUUCCAUUUGA
494	AAACAAAUGCAUAAUCU
495	AUAUUUCUUUCCAUUUG
496	UAUAUUUCUUUCCAUUU
497	UUAUAUUUCUUUCCAUU
559	CAGUCACUCUAAGUUACUU U
560	UGAAUUCAAUCUUCAACCCU
561	AAGAUUGAAUUCAUAACUA U
562	UUAAAUAAAGCAUAGUGCA A
563	UGUGGUUUUUAAAUAAAGC A
564	UAUUUAUGAGAUCAACAGC A
565	AUCUCAUAAAUAGAACUUG U
566	CUGUUGAUAGACACUAAAA G
567	AUAGACACUAAAAGAGUAU U
568	GAAUAUAAGGCUAUAAAUA U
569	AAUUAUUUCUUCUGCCUUU A
570	AAAGGCAGAAGAAAUAAUU G
571	CAAAAUAUUGGGCUCUGAU U
572	UUCUGUAGGAAUCAGAGCC C
573	CUUUUUAUAUUUCUUUCCA U

Seq ID No:	Spacer sequence (5' → 3'), shown as RNA
346	UAUUAAUAAGAUAACCU
347	UGUAUUUGUGAAGUCUU
348	GUAUUUGUGAAGUCUUA
349	UAUUUGUGAAGUCUUAC
350	AAUUUUAUUAAUAAGAU
351	GAAUUUUAUUAAUAAGA
352	GUGAAGUCUUACAAGGU
353	UGAAGUCUUACAAGGUU
354	CCUAGGAUGUUUGAAUU
355	CAAGGUUAUCUUAUUAA
356	ACCUAGGAUGUUUGAAU
357	UACCUAGGAUGUUUGAA
358	UUACCUAGGAUGUUUGA
359	UUUACCUAGGAUGUUUG
360	UUUUACCUAGGAUGUUU

Seq ID No:	Spacer sequence (5' → 3'), shown as RNA
422	UCAUUUUAGUCUGUCUU
423	CAUUUUAGUCUGUCUUU
424	GUGUCUAUCAACAGCAA
425	AUUUUAGUCUGUCUUCU
426	AGUGUCUAUCAACAGCA
427	UAGUGUCUAUCAACAGC
428	UAGUCUGUCUUCUUGGU
429	AGUCUGUCUUCUUGGUU
430	GUCUGUCUUCUUGGUUG
431	UUGGUUGCUGUUGAUAG
432	GUUGCUGUUGAUAGACA
433	GAUAAUAUCUAAUACUC
434	CUGUUGAUAGACACUAA
435	AAACUUAGAUAAUAUCU
436	AUAGACACUAAAAGAGU

Seq ID No:	Spacer sequence (5' → 3'), shown as RNA
498	CUUUUUAUAUUUCUUUC
499	AAUACUGAAGAAAACAA
500	UUAAAUACUGAAGAAAA
501	UUACUUCUUGUUUUCUU
502	CUUCUUGUUUUCUUCAG
503	UUGUUUUCUUCAGUAUU
504	UUUUCUUCAGUAUUUAA
505	UCUUCAGUAUUUAACAA
506	CUUCAGUAUUUAACAAU
507	UUCAGUAUUUAACAAUC
508	AGUAUUUAACAAUCCUU
509	AACAAUCCUUUUUUUUC
510	ACAAUCCUUUUUUUUCU
511	CUGGGAAAAUAGAAUAAA AG
512	UUUUACUGGGAAAAUAG AU

Seq ID No:	Spacer sequence (5' → 3'), shown as RNA
574	AAUACUGAAGAAAACAAGA A
575	UCUUCAGUAUUUAACAAUC C
576	CAUCUUUAAAGAAUUAUUU U
577	CCAUUUUAGAAAUAAAUGC C
578	GCAUUUAUUUCUAAAAUGG C
579	UAUUUGUGAAGUCUUACAA G
580	AAUUUUAUUAAUAAGAUAA C
581	UGAAGUCUUACAAGGUUAU C
582	AAGUCUUACAAGGUUAUCU U
583	ACCUUUUUUUUUUUUUACC U
584	CACUUUCCUUAGUGCGCAAA
585	CGCACUAAGGAAAGUGCAA A
586	AAGUUUCAGUCACUCUAAG U
587	UGAAGUUUCAGUCACUCUA A
588	AAUUCAAUCUUCAACCCUAU

Seq ID No:	Spacer sequence (5' → 3'), shown as RNA
361	UUUUUACCUAGGAUGU U
362	UUUUUUACCUAGGAUG U
363	UUUUUUUACCUAGGAU G
364	UCUUAUUAAUAAAAU UC
365	UUUUUUUUACCUAGGA U
366	UUUUUUUUUACCUAGG A
367	UUUUUUUUUUACCUAG G
368	UUAAUAAAAUUCAAAC A
369	AUAAAAUUCAAACAUC C
370	UGACCUUUUUUUUUUU U
371	AAACAUCCUAGGUAAA A
372	CAGUCACUAAACAAUU C
373	GUGCGCAAAAGAAAAU U
374	CUUAGUGCGCAAAAGA A
375	CCUUAGUGCGCAAAAG A

Seq ID No:	Spacer sequence (5' → 3'), shown as RNA
437	UAUUCAAACUUAGAU AA
438	UAGCCUUAUAUUCAAA C
439	AUAGCCUUAUAUUCAA A
440	AAUAUUUAUAGCCUU AU
441	GAUAUUAUCUAAGUU UG
442	UUAAAUAUUUAUAGC CU
443	UCUAAGUUUGAAUAU AA
444	AAAAUUAUUAAAUAU UU
445	AAAAAUUAUUAAAUA UU
446	UAAAAAUUAUUAAAU AU
447	GAAUAUAAGGCUAUA AA
448	AAUAUAAGGCUAUAA AU
449	CCAAGAAUACUAUUUU A
450	AAUUACCAAGAAUACU A
451	AAUAAUUUUUAAAAU AG

Seq ID No:	Spacer sequence (5' → 3'), shown as RNA
513	CUAAAACUUUAUUUUACUG G
514	UUCUAUUUUUCCCAGUAAAA U
515	AAGAUGCAGAGUUUACUA AA
516	GUAAACUCUGCAUCUUUAA A
517	GAAAUAAAUGCCAAAAUA AU
518	AAGAAUUAUUUUUGGCAUU UA
519	GCAUUUAUUUCUAAAAUG GC
520	UUUCUAAAAUGGCAUAGU AU
521	UUAAUAAGAUAACCUUGU AA
522	UAUUUGUGAAGUCUUACA AG
523	AAUUUUAUUAAUAAGAUA AC
524	UGAAGUCUUACAAGGUUA UC
525	CCUAGGAUGUUUGAAUUU UA
526	CACUUUCCUUAGUGCGCAA A
527	GUGACUGUAAUUUUCUUU UG

Seq ID No:	Spacer sequence (5' → 3'), shown as RNA
589	GGAUAGUUAUGAAUUCAAU C
590	GUUUUUAAAUAAAGCAUAG U
591	UGGUUUUUAAAUAAAGCAU A
592	AGAUCAACAGCACAGGUUU U
593	CUUUAUUUAAAAACCACAA A
594	CUGUUGAUCUCAUAAAUAG A
595	UCUUCUUGGUUGCUGUUGA U
596	AAUAUAAGGCUAUAAAUAU U
597	CCUUUAAACAGAAGAAUAA U
598	UUUAAAGGCAGAAGAAAUA A
599	AUUCCUACAGAAAAACUCA G
600	UUUCAAAAUAUUGGGCUCU G
601	CAUUUGUUUCAAAAUAUUG G
602	UAGGAAUCAGAGGCCCAAUA U
603	ACUUAGAUUAUGCAUUUGU U

Seq ID No:	Spacer sequence (5' → 3'), shown as RNA	Seq ID No:	Spacer sequence (5' → 3'), shown as RNA	Seq ID No:	Spacer sequence (5' → 3'), shown as RNA	Seq ID No:	Spacer sequence (5' → 3'), shown as RNA
376	UUUAGUGACUGUAAUUU	452	AUAAUUUUUAAAAUAGU	528	CGCACUAAGGAAAGUGCAAA	604	AAACAAAUGCAUAAUCUAAG

[681] **TABLE 5** provides exemplary spacer sequences that targets Intron-1 of *SERPINA1*.

TABLE 5: Exemplary Spacer Sequences Targeting Safe Harbor Intron-1 SERPINA1

SEQ ID NO:	Spacer Sequence	SEQ ID NO:	Spacer Sequence	SEQ ID NO:	Spacer Sequence	SEQ ID NO:	Spacer Sequence
911	AAGCCCAGAGAGGGUCAGGG	1131	GUGAAAUAAGUCUCAACACC	1351	GCGGGGCCAGGGACAGGGAA	1571	UUGGCAUUUUAUGUGUAUUU
912	AAUUUGCAACACCUGCCAAC	1132	GUGAGUUAUUUAAUGCUCUU	1352	GCUACAUUGAACCCAGAUGC	1572	UUGGUUAGAAUGGCAUCUGG
913	ACCAUAGGCGAGAUUCUCUU	1133	GUGGCCAGGGAAACUGCUGC	1353	GCUAUUCUUCUGAUUCUUUC	1573	UUGUAACAAUCCCGUGAGGU
914	ACCAUUGCUCGGCAGAGCCU	1134	GUGUCCCCACCCUGCAGAGG	1354	GGACAGGGAUGAGGAAUAAC	1574	UUGUGCCUCAUUUUCUUCUU
915	ACCUGUCUCCAAAACUCAGU	1135	GUUCUAACAACACUAGCUAG	1355	GGAGACAGGUCAAAAAGAUG	1575	UUGUGUCUGGGACCACAGAG
916	ACUUCUGUCCCCUCCCCUCA	1136	UAGAUAGAAAAAUUGAGCCC	1356	GGAGUCCUGAUGUUAUAAGG	1576	UUUCAGUGUUACUGAUGUCG
917	AGACAGAGUCUCGCUUUAUC	1137	UCUGCCAGGGUGGAGUGCAG	1357	GGAGUGAAAUCUGUUUAAUC	1577	UUUCUGUUUGCACAGCUCCU
918	AGCACUAGCUUCAUAGCUGU	1138	UGCUCCCGGGUAGAAAUGGU	1358	GGAUUACUUAGAGGCACGAG	1578	UUUGAGCACUAGCUUCAUAG
919	AGCCCAGAAUUCCCCUGUCU	1139	UGUAUGAGAAUUUAGUGAAA	1359	GGGAAAUGAAAAACUGGAGA	1579	UUUUACAGAUGAAGAAACCA
920	AGGACAGAUUAGAUGAUACG	1140	UGUAUGAGGCAAAAUGAGGA	1360	GGGAAGCUGAGACAGAAGAA	1580	UUUUCUCCAUUGUACAGCUA

SEQ ID NO:	Spacer Sequence
921	AGGGAGGGGUUGGCAG GUGU
922	AUGUAAGGCUGAGGCC UGGC
923	AUGUAAGGUUUGGUGU UGAG
924	CAACACCUGCCAACCCC UCC
925	CACAACUGGGGAGAGU GAGG
926	CACAGCUCCCUCUGUCU GUCG
927	CACCUUAUAGGGUAAG CUUU
928	CACUUCAUUUGUUGCU UUAU
929	CAUCUUAGGGUCAUUG CAGA
930	CCAAACUCAGAUGCUG UUCA
931	CCCAAGAGAACAGAGA GGUU
932	CCUCAUACAUAAAGUG AUGU
933	CGGGGCCAGGGACAGG GAAG
934	CUACAUUGAACCCAGA UGCC
935	CUAUUCUUCUGAUUCU UUCA

SEQ ID NO:	Spacer Sequence
1141	UUGCCCAGUUCAACAG GUUG
1142	UUGGAAGGAAACUUUU AAAA
1143	UUUAUUGCCCAGUUCA ACAG
1144	AACAGAUUUCACUCCA AAAC
1145	AACCAGCCUGUGCCAA GUAC
1146	AACCUCCCUGGGAGCUG UUCC
1147	AAGGCUCUGAGGCCAG AUUG
1148	AAGGGAGAUGGAAAUA UACA
1149	AAUAACUCACUAAAAA UCAC
1150	AAUACACAUAAAAAUGC CAAG
1151	AAUCCUUCCAACACUU CAGA
1152	AAUUAUCCAACUGUUU ACUG
1153	ACAAAAUACAAAGAAA AAGG
1154	ACACAUGUGAGCACGG AGAA
1155	ACCUCCUCAUCUAUAG AAGG

SEQ ID NO:	Spacer Sequence
1361	GGGAGGCCAAGGCAGG CGGA
1362	GGGGCCAGUUAUUCAC CUCU
1363	GGGGCUCAAUUUUUCU AUCU
1364	GGGUUUUUUUUGAGAC AGAG
1365	GGUUUAAAGCUUGGCC UCAG
1366	GGUUUCUUCAUCUGUA AAAG
1367	GGUUUUGUGACUAGUA GUUA
1368	GUAAAACAAUGGCUGU ACCA
1369	GUAAGCAAGUAGCUCU GUGA
1370	GUACACAGAGAUGAUU CAAU
1371	GUAUUUUGUUAAGAUG AUGG
1372	GUAUUUUUAGUAGAGC UGAG
1373	GUCACUGACUUGCUGU AUAG
1374	GUCAUAGAGCUAUUGU GAAA
1375	GUCUGGGAUUAAACAG AUUU

SEQ ID NO:	Spacer Sequence
1581	UUUUGGUUUAAAGCUU GGCC
1582	UUUUGUAAAACAAUGG CUGU
1583	AACAGGUUGUUUGACCA UAG
1584	AAGAAAUUCUUCUGUCU CAG
1585	AAGACCUGCCUGGCCAA CAC
1586	AAGAGUUUUGGGGACUG GAGC
1587	AAUGUAGCAAAAACAGA UGGA
1588	ACUGCUGAGCAGGAGCC CCU
1589	AGAGAGGUUAAGUAAC UUGU
1590	AGCCACAGGCCCCACCA CUC
1591	AGUUGGGAAGUGAAAG AGAA
1592	AUGAUCUGAGUGCUGG GAAA
1593	AUUAUGCCUCCCAAGCU GUU
1594	CAGCUUUGUCACUGACU UGC
1595	CAUUUGGAUUACUUAG AGGC

SEQ ID NO:	Spacer Sequence
936	CUUAGCAUAGGGUUAG GCAG
937	GACAGGGAUGAGGAAU AACU
938	GAGACAGGUCAAAAAG AUGA
939	GAGUCCUGAUGUUAUA AGGA
940	GAGUGAAAUCUGUUUA AUCC
941	GAUUACUUAGAGGCAC GAGG
942	GGAAAUGAAAAACUGG AGAA
943	GGAAGCUGAGACAGAA GAAU
944	GGAGGCCAAGGCAGGC GGAU
945	GGGACUGGAGCCAGGU GCCU
946	GGGCCAGUUAUUCACC UCUC
947	GGGCUCAAUUUUUCUA UCUA
948	GGUUUUUUUUGAGACA GAGU
949	GUGUUGAGACUUAUUU CACU
950	GUUUAAAGCUUGGCCU CAGU

SEQ ID NO:	Spacer Sequence
1156	ACCUCUCUGAACCUCA GUUU
1157	AGCCUCAUAACCCUCC AAGG
1158	AUCACGCACUGAGCUU AGCC
1159	CAAAACUGAGAUGUAA UCCU
1160	CAGGAACCAGUGUAUC CACC
1161	CAGGCGUGUGCCACCA CGCC
1162	CAGGGCUCAGUGGUCA GGAG
1163	CAUCAAAAGGAUCUGA GAGG
1164	CAUCAAAAUGGGCCUC CCCA
1165	CAUCUCAGUUUUGUAA GCAA
1166	CAUCUGCAAUGACCCU AAGA
1167	CCAUAUGCCAAGCACU GUUC
1168	CCCUAUAAGGUGCAAA UUAU
1169	CCUGCCGUCUGGGUCA CUAA
1170	CGUCAGUUAUUCCUCA UCCC

SEQ ID NO:	Spacer Sequence
1376	GUGACUAGUAGUUAAC AAUG
1377	GUGCAGUCCUUCAUGG CUUA
1378	GUUAAGAUGAUGGUUU UGGU
1379	GUUAUUUCCAGCUUUU UUUC
1380	GUUGCUGUUGCUGUAU CUUG
1381	GUUGCUUUAUUGGAAG GAAA
1382	UAAAAGUCCAAGGGGG UUGA
1383	UAAAAGUUUCCUUCCA AUAA
1384	UAACUUCAGAUAGACC UGGG
1385	UACAGAUGAAGAAACC AAAG
1386	UACAGAUGAGGAAGAU GGGG
1387	UAGAGGCGUUCUAUAG GAAA
1388	UAGCACCGUCUUUAGU UCUA
1389	UAGCCACAGACUGAUG CUCC
1390	UAGGAGUGCAGGGUGC UGCC

SEQ ID NO:	Spacer Sequence
1596	CCACCCAACACUCACAA GGA
1597	CCCAUCCAGCGCCUGAG ACC
1598	CUCCCCACCCCUCUCUG GAC
1599	CUGGUCAUAUCUGGAGG GGA
1600	CUGUAAGCAGAAGUGG AUGA
1601	CUUAUAACAUCAGGACU CCA
1602	UAACAACACUAGCUAGC UGA
1603	UACCUAGGGGGGCUGCAG GCU
1604	UAUAGGAAACUGGCCAC CUU
1605	UCCGUGCUCACAUGUGU UAA
1606	UCGAGAAUACCAGUUCU UGC
1607	UCUUGGGCAAAACAUGA UAA
1608	UGCUACUUCCUCAGUGA CAU
1609	UGGAGAGGCUUCUGAGC UCC
1610	UGGGCUCUUGCUGUCAG GGG

SEQ ID NO:	Spacer Sequence
951	GUUUCUUCAUCUGUAA AAGA
952	GUUUUGUGACUAGUAG UUAA
953	UAAAACAAUGGCUGUA CCAU
954	UAAGCAAGUAGCUCUG UGAU
955	UACACAGAGAUGAUUC AAUG
956	UAUUUUGUUAAGAUGA UGGU
957	UAUUUUUAGUAGAGCU GAGG
958	UAUUUUUCUUCAGCUA GCUA
959	UCACUGACUUGCUGUA UAGC
960	UCAUAGAGCUAUUGUG AAAU
961	UCUGGGAUUAAACAGA UUUC
962	UGACUAGUAGUUAACA AUGU
963	UGCAGUCCUUCAUGGC UUAU
964	UGCUGCUUUAAAGUAG AUGG
965	UUAAGAUGAUGGUUUU GGUU

SEQ ID NO:	Spacer Sequence
1171	CUAACACCUGGUGUUU UACA
1172	CUGGCAGGGGGCCUGG GGUA
1173	CUUAGAGGCACGAGGG CCAG
1174	GAGGCACGAGGGCCAG GAGA
1175	GAUAGAAGCCCUCCAA GGGC
1176	GAUCCCAGGUCUAUCU GAAG
1177	GAUGAUACGGUGCCUG UAAA
1178	GCACCUCACGGGAUUG UUAC
1179	GCAUAGGGUUAGGCAG CUGG
1180	GCAUUAGCACCUCACG GGAU
1181	GCCAGGCGUGGUGGCA CACG
1182	GCCCUAUUUUAUAGAU AGAA
1183	GCCGUGGCCUGGUGCA GAGC
1184	GCUGUGUGACCUUGGA CAAG
1185	GGCAGGGAGUGGGCUCG CUUC

SEQ ID NO:	Spacer Sequence
1391	UAGUAGAGCUGAGGUU UCAC
1392	UAGUGAGUUAUUUAAU GCUC
1393	UAUAGAUAGAAAAAUU GAGC
1394	UAUGCUCCCGGGUAGA AAUG
1395	UCACUUCCCACCUCCUC CAU
1396	UCAUAAUAACAGCAGC CAUG
1397	UCAUUUCCCAAAUUCA AGGA
1398	UCCAUUAAAUUAUCCA ACUG
1399	UCCCAAGAGACGGGGC UGUC
1400	UCCUGCACUGAGUUUU GGAG
1401	UCUAUCUAUAAAAUAG GGCU
1402	UCUCCAGUUUUUCAUU UCCC
1403	UCUCCAUUGUACAGCU AUGA
1404	UCUUCAGCUAGCUAGU GUUG
1405	UCUUCUUUUGUAAAAC AAUG

SEQ ID NO:	Spacer Sequence
1611	UGUGUUGCUGUGGAUG ACGU
1612	UUGCUGCUAUUGCUGCA GGU
1613	UACACUAACAUAGGUUU UUG
1614	UCACUUUCCAGGUGUAU GAG
1615	UCCUCAAAGCUUACCCU AUA
1616	GACAGCAACACUUACAU CUG
1617	AGGCCAGAUUGCUCCAC GUC
1618	GGAUCGGGGCAGGUGG GAGC
1619	AGAGGGCAACUCUUUUG AGG
1620	AGUUUGGCAAAUUUUCC AUU
1621	CCGAGCAAUGGUCAAAG UCA
1622	GAGGGGAUGGAGAAUG UGAG
1623	GGUCACUAAUGGUGUCC CAG
1624	CCAGGGUGGAGUGCAGU GGC
1625	GUUUCCUCCUUUCACAG UGA

SEQ ID NO:	Spacer Sequence
966	UUAUUUCCAGCUUUUU UUCU
967	UUGCUGUUGCUGUAUC UUGG
968	UUGCUUUAUUGGAAGG AAAC
969	AACUGGGCAAUAAAUA AACC
970	ACGUGGAGCAAUCUGG CCUC
971	ACUUUGACCAUUGCUC GGCA
972	AGACUUAUUUCACUAA AUUC
973	AGCAACUGUCUGAAGU CACC
974	AGCUCCUGCCCGAGAG AUGG
975	AGGCUAUACAGCAAGU CAGU
976	CAAAUUCAAAGGCUUC AGGU
977	CCAUUGUGCAAGAUGC AUUU
978	CCCUCUCAGAUCCUUU UGAU
979	CUCAACCUCUCUGUUC UCUU
980	CUCUUGGUCCCUUAUU UUAU

SEQ ID NO:	Spacer Sequence
1186	GGGUCAUUGCAGAUGU AAGU
1187	GGGUUUUUUUCCCAAG AGAC
1188	GGUCAGCUAUUCAAUG UCAC
1189	GUGACCCAGACGGCAG GUAA
1190	GUGCUGUAUAUUUCCA UCUC
1191	UAACAUCAGGACUCCA AAAC
1192	UAGGAGACAACAGACA GGAG
1193	UAGGGUAAGCUUUGAG GACA
1194	UCCAACUGUUUACUGC CUGC
1195	UCCAGGAAUCGGGCCA GAUC
1196	UGAAAAUAGGAGCUCA GCUG
1197	UGCCUCCCAAGCUGUU CCUU
1198	UGGAAAUUCCCCCAUC AUUA
1199	UGGAUGGUGUGUGAUU UGGG
1200	UGUAAUGUGUGGUCUG UUCC

SEQ ID NO:	Spacer Sequence
1406	UGACCUGUCUCCAAAA CUCA
1407	UGAGACAGAGUCUCGC UUUA
1408	UGAGGGAGGGGUUGGC AGGU
1409	UGAUGUAAGGCUGAGG CCUG
1410	UGAUGUAAGGUUUGGU GUUG
1411	UGCCUCAUACAUAAAG UGAU
1412	UGCUAUUCUUCUGAUU CUUU
1413	UGGGUUUUUUUUGAGA CAGA
1414	UGGUUUAAAGCUUGGC CUCA
1415	UGUAAAACAAUGGCUG UACC
1416	UGUAUUUUUAGUAGAG CUGA
1417	UGUUAAGAUGAUGGUU UUGG
1418	UUAGUAGAGCUGAGGU UUCA
1419	UUAGUGAGUUAUUUAA UGCU
1420	UUCCCAAGAGACGGGG CUGU

SEQ ID NO:	Spacer Sequence
1626	CUUACAGGAACCAGUGU AUC
1627	UAAAAUGGGAAUCACAC UAG
1628	GGUUCAAUGUAGCAAA ACAG
1629	GACCUUAACCUCCUCAU CUA
1630	GGCUCAAAGGAGGUGG AGAG
1631	UGGUCCCAGACACAAGA AUA
1632	GUCCUGGGGCCUCACUC UCC
1633	GCCAAGUACCCCAUGGU GUA
1634	UCCCCAUCUUCCUCAUC UGU
1635	UUCCUGGUCAUAUCUGG AGG
1636	CAGGGUGGGGACACUAA AUA
1637	UGGCUAAAAUUAAAGG CUCU
1638	GGCUCUUGCUGUCAGGG GCU
1639	GGAUUAAACAGAUUUC ACUC
1640	UGUACAAAGCUGAAGCC AGA

SEQ ID NO:	Spacer Sequence
981	CUGUAUCUUGGCUGGU GUCC
982	CUGUCCAGAGCCUGGG UGCA
983	CUGUGGAUGACGUUGA CUUU
984	CUGUUGCUGUAUCUUG GCUG
985	CUUUAUUGGAAGGAAA CUUU
986	GAAGGAUUUAAUUUUU CCUG
987	GCAGGUGUUGCAAAUU CAAA
988	GCCAGGCAGGUCUUGA ACUC
989	GGAAGUGAAAGAGAAA UUAA
990	GGCCUCUGCAGGGUGG GGAC
991	GGUGGGAACAGCUCCC AGGA
992	GGUUCAUUAUGCCUCC CAAG
993	UCUCCUAUAAGCCUCA CCAC
994	UGCAAAGUGGGGCUGG GAGA
995	UGGAGUGGGUAUCCGC CUGC

SEQ ID NO:	Spacer Sequence
1201	UUAAACCAGCCUGUGC CAAG
1202	UUAGCCCUAUUUUAUA GAUA
1203	UUCAGGAAUGCCAGCU UGAG
1204	UUUCACUAAAUUCUCA UACA
1205	UUUUAUGCUCCCGGGU AGAA
1206	AAAGCUUGGCCUCAGU GUCC
1207	AAUAAUCCCUAGGCAC UUCC
1208	AAUCACAGCCCCUCCA UGCU
1209	AAUCCCAGACAAAACA UUUA
1210	ACUGCCUGCCAGCUGC CUAA
1211	AUGUAUGAGAAUUUAG UGAA
1212	AUUUAUUGCCCAGUUC AACA
1213	CACCGUGUUGGCCAGG CAGG
1214	CAGUGAGCCGAGAUCA CGCC
1215	CAUACCCGGGGCUGAG ACAA

SEQ ID NO:	Spacer Sequence
1421	UUCCUGCACUGAGUUU UGGA
1422	UUCUAUCUAUAAAAUA GGGC
1423	UUCUCCAGUUUUUUCAU UUCC
1424	UUCUUCAGCUAGCUAG UGUU
1425	UUGACCUGUCUCCCAAA ACUC
1426	UUGAGACAGAGUCUCG CUUU
1427	UUGUAUUUUUAGUAGA GCUG
1428	UUUCCCAAGAGACGGG GCUG
1429	UUUCUCCAGUUUUUCA UUUC
1430	UUUGAGACAGAGUCUC GCUU
1431	UUUUCUCCAGUUUUUC AUUU
1432	UUUUGAGACAGAGUCU CGCU
1433	AAACUCUUCACUGUGA AAGG
1434	AAAGGCUUCAGGUGAU AGGC
1435	AACCCCCUUGGACUUU UAAA

SEQ ID NO:	Spacer Sequence
1641	ACUCUGUCACUUUCCAG GUG
1642	AAGUCACCGAGCAUCAG AGA
1643	CCCUGCACCUCAGCAGG CGG
1644	AACCUCAGUUUCCUGGA GAU
1645	UCUCAGCUUCCCAAAUA GCU
1646	CAAUGACCCUAAGAUGC AAA
1647	GCUUUGGUUUCUUCAUC UGU
1648	GGGCAGGAGGGGGGUU GUGG
1649	UCUGUCGGGGGCUCCUG UCU
1650	CACCAGGCCACGGCUAA GCU
1651	AAGUUUUCUGAUUCAA GACU
1652	GGCUUCAAACUCUUCAC UGU
1653	AGCUCCCCAUGGCCCAG GCA
1654	CAGUCCUGGCACCCCAG GAU
1655	GCCUCAGAGCCUUUAAU UUU

SEQ ID NO:	Spacer Sequence
996	UGGCUUCUAGCCCUGG CAGG
997	UGUCCUGUUUCCCAGU CCUG
998	UUAGAACUAAAGACGG UGCU
999	UUUGACCAUAGGCGAG AUUC
1000	AAAGCAGGGGUUUGUG CUGC
1001	AACCCAGAUGCCAUUC UAAC
1002	AACCCGGGAGGUGGAG GUUU
1003	AACCUGGCCCUAGUGU CACU
1004	AACUCCUGACCUCUUG AUCC
1005	AAUAGCUGACCUAAUC UCUC
1006	AAUCAGAAAACUUCAG AGGC
1007	AAUCAUCUCUGUGUAC AAAG
1008	AAUUUGGGAAAUGAAA AACU
1009	ACACCACGUGGAAGUG CCUA
1010	ACAUGAUUCUUGAUUC UGAU

SEQ ID NO:	Spacer Sequence
1216	CCCAGUCCUGCCCGGU CCCC
1217	CCCUGGCCACUAAAUA GUGG
1218	CCUAUAGAACGCCUCU AAAA
1219	CCUCCUUUCACAGUGA AGAG
1220	CCUGGAGAUGAGAAAG GUGG
1221	CCUGUUCCCCAUCCAG CGCC
1222	CCUUCCAAUAAAGCAA CAAA
1223	CUCUCCACCACCUUUC UCCU
1224	CUCUUGGGUCUCAGGC GCUG
1225	CUUCAUCUGUAAAAGA AGGA
1226	GAAGCCCAGAGAGGGU CAGG
1227	GACCAUAGGCGAGAUU CUCU
1228	GCACAGCUCCUCUGUC UGUC
1229	GCUUAGCAUAGGGUUA GGCA
1230	GGCAAAUUUUCCAUUA AAUU

SEQ ID NO:	Spacer Sequence
1436	AAGACUUCUGGCUUCA GCUU
1437	AAGGACACCUUUUUCU UUGU
1438	AAGGAGAGAUGAGCGU CUCU
1439	AAUGUCACUGAGGAAG UAGC
1440	AAUGUCAGGUUUUGGA GUGA
1441	ACAAUAGCUCUAUGAC AAAG
1442	ACAGAUGGGAAAACUG AGGU
1443	ACAGUGAAGAGUUUGA AGCC
1444	ACCGUGUUGGCCAGGC AGGU
1445	ACCUCUCUGUCCCCAUC UUC
1446	ACUCCAAAACCUGACA UUGA
1447	ACUGCGAGAGGUUCUG GAGA
1448	ACUGGGGCCCUCUGUA CACU
1449	ACUGUGAAAGGAGGAA ACCA
1450	ACUUCCCAACUGAACU UUGU

SEQ ID NO:	Spacer Sequence
1656	AUUCAAGACUUCUGGCU UCA
1657	AUUCUUUCAGUGUUACU GAU
1658	UCCCCUCCCCUCACCCC ACC
1659	UUGUCUCCUAUAAGCCU CAC
1660	UCUCAAAAAAAACCCAA AAA
1661	GCCCUGAGGCCACUCAG CAU
1662	UUCUCUUGGGCAAAACA UGA
1663	GAGAGGCUUCUGAGCUC CCC
1664	GUUCUGCUACUUCCUCA GUG
1665	UUUGCACAGCUCCUCUG UCU
1666	AUGCUCGGUGACUUCAG ACA
1667	UGAAGACGGCAGGUUCU ACC
1668	UUUAAUCCCAGACAAAA CAU
1669	UUCCAUUUGGAUUACUU AGA
1670	AUCUGGCCCGAUUCCUG GAU

SEQ ID NO:	Spacer Sequence
1011	ACUCGGGGGCCUGGCCC CCCC
1012	AGACCAAUGUCACAGC CCAA
1013	AGAUUCUUCAACCCCC UUGG
1014	AGCCCCAAAGAGCAUU AAAU
1015	AGCCCUGGGUGACAGU GCCC
1016	AGGCCUCUCUGAUGCU CGGU
1017	AUCCGCCUGCCUUGGC CUCC
1018	AUCUCUGAAGUGUUGG AAGG
1019	AUUCUGAUCUGGCCCG AUUC
1020	CACAAUGGCAACAGCU AGAG
1021	CAGAUGUAAGUGUUGC UGUC
1022	CCACACUUACUAACAC CUGG
1023	CCCAGUUCAACAGGUU GUUU
1024	CCGACAUCAGUAACAC UGAA
1025	CUAAAGGCCAAGUGGA UAGU

SEQ ID NO:	Spacer Sequence
1231	GGGGACUGGAGCCAGG UGCC
1232	GGUGUUGAGACUUAUU UCAC
1233	GUAUUUUUCUUCAGCU AGCU
1234	GUGCUGCUUUAAAGUA GAUG
1235	UACAAAAGAAGAAAAU GAGG
1236	UACAGGCACCGUAUCA UCUA
1237	UCCCAUCUGUGAAGAC GGCA
1238	UCUGAUUCAAGACUUC UGGC
1239	UGCCCAAGAGAACAGA GAGG
1240	UGCUACAUUGAACCCA GAUG
1241	UGGAGACAGGUCAAAA AGAU
1242	UGGAGUCCUGAUGUUA UAAG
1243	UGGAGUGAAAUCUGUU UAAU
1244	UGGUUUUGUGACUAGU AGUU
1245	UGUAAGCAAGUAGCUC UGUG

SEQ ID NO:	Spacer Sequence
1451	ACUUCCCACCUCCUCCA UCU
1452	ACUUCUUUGAGCACUA GCUU
1453	AGACAGUUGCUCAACC UCUC
1454	AGAGAUCAAGGUCGGG UGGG
1455	AGAGGCAGCCUGGCUA AAUG
1456	AGAUAGACCUGGGAUC UAAU
1457	AGCAGGAAGGGGCUCC UGCU
1458	AGCUAGCUAGUGUUGU UAGA
1459	AGCUUUGUACACAGAG AUGA
1460	AGGAAUGCCAGCUUGA GCCC
1461	AGGUGAUAGGCCAGUG AGGA
1462	AGUGAGCCGAGAUCAC GCCA
1463	AGUGUUACUGAUGUCG GCAA
1464	AUAAUAACAGCAGCCA UGAG
1465	AUACCCGGGGCUGAGA CAAU

SEQ ID NO:	Spacer Sequence
1671	ACAUUCAUCAUCUGAGA UCU
1672	UAAAAGAAGGAUAUUA GCAU
1673	UGUUGCUGUGGAUGAC GUUG
1674	GGACCACAGAGCAUUGU GGG
1675	GCCCGAUUCCUGGAUAA UCG
1676	CUACUUCCUCAGUGACA UUG
1677	GCUUCAGCUUUGUACAC AGA
1678	UCGGGGGCUCCUGUCUG UUG
1679	AGGACACAGCUGUCUCA GUU
1680	AGUGCUGGGAAAGCCAA GAC
1681	AAGUGUUGGAAGGAUU UAAU
1682	GAGUAGAAGGAUCACU GUGG
1683	AUAGAAGGGCCAGGAG AGAG
1684	UGCAGCAGUUUCCCUGG CCA
1685	GGCCCUCCUGUCCCAGU GGA

SEQ ID NO:	Spacer Sequence
1026	CUAUUAGCUGUGUGUGAC CUUG
1027	CUCCACGUCAACUUGC UAUU
1028	CUCGGCAGAGCCUGCU CUCG
1029	CUGCAGGUCUUUUAGA GGCG
1030	CUGCGCAGCAGGCACU UCUU
1031	CUGCUAUUGCUGCAGG UCUU
1032	CUGUAUAGCCUCAACA AGGC
1033	CUGUCAGGGGGCUUUAG AGAC
1034	CUUACAAAACUGAGAU GUAA
1035	CUUUCUUGUAACAAUC CCGU
1036	GAAGGAAACUUUUAAA AGUC
1037	GACAAGUUACUUAACC UCUC
1038	GACUUUUAAAAGUUUC CUUC
1039	GAGGGCUUCUAUCUAA UCCC
1040	GAGGGUUAUGAGGCUU AAUC

SEQ ID NO:	Spacer Sequence
1246	UGUCUGGGAUUAAACA GAUU
1247	UGUGACUAGUAGUUAA CAAU
1248	UUCAUUUCCCAAAUUC AAGG
1249	UUGGGUUUUUUUUGAG ACAG
1250	UUUUCCCAAGAGACGG GGCU
1251	UUUUUGAGACAGAGUC UCGC
1252	AAAAGUCCAAGGGGGU UGAA
1253	AAAAGUUUCCUUCCAA UAAA
1254	AAACCAAAAGAAUGAA GGCU
1255	AAAGCAGCACAAACCC CUGC
1256	AAAGUAGAUGGAGGAG GUGG
1257	AACUUCAGAUAGACCU GGGA
1258	AAUGCUCUUUGGGGCU CAAU
1259	AAUGGAAAAUUUGCCA AACU
1260	AAUUUCACAAUAGCUC UAUG

SEQ ID NO:	Spacer Sequence
1466	AUAGCUGUACAAUGGA GAAA
1467	AUCAUCUGAGAUCUGA GGAC
1468	AUCUCCCUGGUCCUCA UUUU
1469	AUCUGUAAAAGAAGGA UAUU
1470	AUGGCUUAUGGAAAUU CCCC
1471	AUGGUGACAUUUCUGG GAUC
1472	AUUCUUUUGGUUUAAA GCUU
1473	AUUGCUUUCUUGUAAC AAUC
1474	AUUUAGCCAGGCUGCC UCUG
1475	AUUUCCCAAAUUCAAG GACA
1476	AUUUGUUGCUUUAUUG GAAG
1477	CAACACUUCAGAGAUC AAGG
1478	CAAUAAAGCAACAAAU GAAG
1479	CACGUGGUGUCAAUCC CUGA
1480	CAGCUUGACAUGAUUC UUGA

SEQ ID NO:	Spacer Sequence
1686	UUUUGCUACAUUGAACC CAG
1687	AGAUCUGAGGACACAGC UGU
1688	AAGUUAAAAGAGGUGU CUUU
1689	AAAAUAGGAGCUCAGCU GCA
1690	CCUCCCAAGCUGUUCCU UAU
1691	CUAAGCAAACAUCAGCC UCA
1692	AGAAUUUAGUGAAAUA AGUC
1693	AGGCUUAAUCACGCACU GAG
1694	CCCGGCAGCCACAUUGU UAA
1695	CCCAGGCUUCGGACCAC CAG
1696	CUGAGUGGCCUCAGGGC CAG
1697	GAUGGUGUGUGAUUUG GGGC
1698	GUCAAACAACCUGUUGA ACU
1699	CCAAGCACUGUUCUCCG UGC
1700	AGCCAGGACAAGUCAUU CAU

SEQ ID NO:	Spacer Sequence
1041	GCACAGGCUGGUUUAA UAAU
1042	GCAUAUGGUAAGCACU ACGU
1043	GCCAGAUGGCGACUUU CUCC
1044	GCCUCAGUGUCCGUAC ACCA
1045	GCCUCCCAAAGUGCUG GGGU
1046	GCCUUGGUCAUCUUUU UGAC
1047	GCCUUUAGCAAGAAUC CUCU
1048	GCUGGUGUCCCCCAUC CUGG
1049	GCUUUCCCAGCACUCA GAUC
1050	GGAAAAAAACCCUAAG GGCU
1051	GGAGGCAUAAUGAACC CAAC
1052	GGCAAAACAUGAUAAC CUUU
1053	GGCAUGCAUCUUUCUC CCCU
1054	GGCUGUGACAUUGGUC UCAA
1055	GGUCUCAGGCGCUGGA UGGG

SEQ ID NO:	Spacer Sequence
1261	AAUUUCUCUUUCACUU CCCA
1262	AAUUUUAGCCACAGAC UGAU
1263	AAUUUUUCCUGCACUG AGUU
1264	ACAAAAGAAGAAAAUG AGGC
1265	ACAGAUGAAGAAACCA AAGC
1266	ACAGAUGAGGAAGAUG GGGA
1267	ACAGGCACCGUAUCAU CUAA
1268	ACCCCAGGCCCCCUGC CAGU
1269	AGAGACGCUCAUCUCU CCUU
1270	AGAGGCGUUCUAUAGG AAAC
1271	AGCAAGAAUCCUCUGG UCCU
1272	AGCACCGUCUUUAGUU CUAA
1273	AGCCACAGACUGAUGC UCCC
1274	AGCCAGGCUGCCUCUG AAGU
1275	AGGAGUGCAGGGUGCU GCCU

SEQ ID NO:	Spacer Sequence
1481	CAGCUUUUUUUCUCCA GUUU
1482	CAGGUGUAUGAGCCAG GACA
1483	CAUAAGCCAUGAAGGA CUGC
1484	CAUCUCCCUUUAAUUU CUCU
1485	CAUUAAAUUAUCCAAC UGUU
1486	CCAAAUAGCUGGGAUU ACAG
1487	CCAAAUUCAAGGACAC CUUU
1488	CCAACUGAACUUUGUG CAGU
1489	CCAAGAGACGGGGCUG UCCC
1490	CCACCUCCUCCAUCUAC UUU
1491	CCAGCACUCAGAUCAU GAAC
1492	CCAGUCCUGCCCGGUCC CCC
1493	CCAUCUGUGAAGACGG CAGG
1494	CCAUGGCUAGGCAUGA GACA
1495	CCAUUUUACAGAUGAG GAAG

SEQ ID NO:	Spacer Sequence
1701	GAAAUUCCCCCAUCAUU ACA
1702	GUAAGCACUACGUACAU UUU
1703	AAGCUAGUGCUCAAAGA AGU
1704	ACCAGGAACAGACCUGU GCC
1705	UUAGUGUACAGAGGGCC CCA
1706	UAAUGUGUGGUCUGUU CCAU
1707	UAUGAGGCAAAAUGAG GACC
1708	AAACAUGUAUUCCUCAC UGG
1709	CUCCCGGGUAGAAAUGG UGU
1710	CCUUUCACUGCGAGAGG UUC
1711	ACAAAGUCCACCUUUCU CAU
1712	AGGCAAAAUGAGGACCA GGG
1713	CAGUCCUUCAUGGCUUA UGG
1714	ACCUUGGACAAGUUACU UAA
1715	GUCCCAGACACAAGAAU AGG

SEQ ID NO:	Spacer Sequence
1056	GGUUUAUUUAUUGCCC AGUU
1057	GUCAUCUUUUUGACCU GUCU
1058	GUCCCUUAUUUUAUGC UCCC
1059	GUCUCAAGCAUCAUGU UGGG
1060	GUUAGAAUGGCAUCUG GGUU
1061	UAACAAUCCCGUGAGG UGCU
1062	UACAGCUAUGAAGCUA GUGC
1063	UCCAAGGUCACACAGC UAAU
1064	UCCUGGCUCAUACACC UGGA
1065	UCUCAGCCCCGGGUAU GAAA
1066	UGAGUGUUGGGUGGGA ACAG
1067	UGCAAGAUGCAUUUCC CUUC
1068	UGCCUCAUUUUCUUCU UUUG
1069	UGGGGACAGCCCCGUC UCUU
1070	UGUCUGGGACCACAGA GCAU

SEQ ID NO:	Spacer Sequence
1276	AGGAUUACAUCUCAGU UUUG
1277	AGUAGAGCUGAGGUUU CACC
1278	AGUGAAAUAAGUCUCA ACAC
1279	AGUGAGUUAUUUAAUG CUCU
1280	AGUGGCCAGGGAAACU GCUG
1281	AGUGUCCCCACCCUGC AGAG
1282	AGUUCUAACAACACUA GCUA
1283	AUAGAUAGAAAAAUUG AGCC
1284	AUCUGCCAGGGUGGAG UGCA
1285	AUGCUCCCGGGUAGAA AUGG
1286	AUGUAUGAGGCAAAAU GAGG
1287	AUUGCCCAGUUCAACA GGUU
1288	AUUGGAAGGAAACUUU UAAA
1289	CAAGGAGAGAUGAGCG UCUC
1290	CACAAUAGCUCUAUGA CAAA

SEQ ID NO:	Spacer Sequence
1496	CCCCACACCCAUCCAU UUU
1497	CCCCAUCAUUACAGGG CUCA
1498	CCCUGUCUAGGAGGGC ACCU
1499	CCGGAGAGCCUGCUGC ACUC
1500	CCUGCACCCUGCCGGU UGC
1501	CCUGGCCACUAAAUAG UGGG
1502	CCUGUCCCUGGCCCCGC AAA
1503	CCUUCUGUGCAGCAGU UUCC
1504	CUAUAGAACGCCUCUA AAAG
1505	CUCACCUGGUGCCCACU CAU
1506	CUCACUGGCCUAUCACC UGA
1507	CUCAGUGACAUUGAAU AGCU
1508	CUCAUCCCUGUCCAAA GGAU
1509	CUCAUCUGUAAAAAUGG GAAU
1510	CUCCUUUCACAGUGAA GAGU

SEQ ID NO:	Spacer Sequence
1716	UCCUCAGAUCUCAGAUG AUG
1717	CCAAGUACUUGCCGACA UCA
1718	UUAAUUCAUUGCUUUCU UGU
1719	AUUAAACCUCCUGGGAG CUG
1720	GCUAAAAUUAAAGGCUC UGA
1721	CCACCACGCCUGGCUAA UUU
1722	GCUGAACCAGCGAGAGC AGG
1723	UACAAAGCUGAAGCCAG AAG
1724	AGCCGGCGUUCCUCCCC ACC
1725	ACCAGGCCCUGUCACUU ACU
1726	AUUUGGGGCCAGUUAU UCAC
1727	UGAUUUGGGGCCAGUU AUUC
1728	UGACCUUGGACAAGUUA CUU
1729	CUGCUUUAAAGUAGAU GGAG
1730	CAAAGUGGGGCUGGGA GAAG

SEQ ID NO:	Spacer Sequence
1071	UUAACUACUAGUCACA AAAC
1072	UUACAAGAAAGCAAUG AAUU
1073	UUCCAGCUUUGUCACU GACU
1074	UUGAGGCUAUACAGCA AGUC
1075	AAAGAGGUGUCUUUUC ACUU
1076	ACAAUGUGGCUGCCGG GCAU
1077	ACUACUAGUCACAAAA CCAA
1078	AGAUGAUGGUUUUGGU UUUG
1079	AGGUCCAGAGAGGGGU GGGG
1080	AGUAACUUGUCCAAGG UCAC
1081	AUUCAUUGCUUUCUUG UAAC
1082	CAAGAAAGCAAUGAAU UAAC
1083	CAGGCGAGAGCCCCUG CACC
1084	CUAGUGUGAUUCCCAU UUUA
1085	CUGAUGUCGGCAAGUA CUUG

SEQ ID NO:	Spacer Sequence
1291	CACAGUGAAGAGUUUG AAGC
1292	CACUAAAUUCUCAUAC AUAA
1293	CACUCCAAAACCUGAC AUUG
1294	CACUGCGAGAGGUUCU GGAG
1295	CACUUCCCAACUGAAC UUUG
1296	CACUUCCCACCUCCUC CAUC
1297	CAGCAGGAAGGGGCUC CUGC
1298	CAGUGUUACUGAUGUC GGCA
1299	CAUAAUAACAGCAGCC AUGA
1300	CAUUUCCCAAAUUCAA GGAC
1301	CCAGCUUUUUUUCUCC AGUU
1302	CCAGGUGUAUGAGCCA GGAC
1303	CCAUAAGCCAUGAAGG ACUG
1304	CCAUCUCCCUUUAAUU UCUC
1305	CCAUUAAAUUAUCCAA CUGU

SEQ ID NO:	Spacer Sequence
1511	CUCUUCUUGUGCCUCA UUUU
1512	CUGAAUAAUCGCUCUG CACC
1513	CUGCACUGAGUUUUGG AGAC
1514	CUGCUGAAAUGGUCCU GGGG
1515	CUGGAGAUGAGAAAGG UGGA
1516	CUGGAUAAUCGUGGGC AGGC
1517	CUGGUGUCACAUCACC UACU
1518	CUGUCCUCCACUGGGA CAGG
1519	CUGUUCCCCAUCCAGCG CCU
1520	CUUCCAAUAAAGCAAC AAAU
1521	GGACCACCAGCCUGCA GCCC
1522	GGGCUCAGGCACCUGG CUCC
1523	UAACCAAGAAUCCUGG CUGG
1524	UACCCGGGAGCAUAAA AUAA
1525	UACUCCAGAUAUGCCC AGGC

SEQ ID NO:	Spacer Sequence
1731	CAAACAGAAAGAAAUG GCUG
1732	ACAUUGGUCUCAAGCAU CAU
1733	GCUUCUAGCCCUGGCAG GGC
1734	GGGCAGGGCAGGUUCUG UGU
1735	UGGUCUGUUCCAUUUGG AUU
1736	GUCUGUUCCAUUUGGAU UAC
1737	GGGGAAUGUGCUCCUCC CAG
1738	ACACCAGGAAAGAUGGA UGG
1739	AAGACGGCAGGUUCUAC CUA
1740	GAGUGGGUAUCCGCCUG CUG
1741	UCUGGGACCACAGAGCA UUG
1742	CCUCAUUUUCUUCUUUU GUA
1743	AGUGUUGGGUGGGAAC AGCU
1744	AGCACGGAGAACAGUGC UUG
1745	UCCUGUUUCCCAGUCCU GCC

SEQ ID NO:	Spacer Sequence
1086	CUUAACCUCUCUGAAC CUCA
1087	GAACUAAAGACGGUGC UAAA
1088	GAAUGGCAUCUGGGUU CAAU
1089	GGCAGCUGGCAGGCAG UAAA
1090	GUAAGUGUGGCAAUGA UGGU
1091	GUGUACAGAGGGCCCC AGUG
1092	UAAGGAACAGCUUGGG AGGC
1093	UCAUGUUUUGCCCAAG AGAA
1094	UGAGGCUUAAUCACGC ACUG
1095	UUAUGAAAAUAGGAGC UCAG
1096	UUCACCUCUCUGUCCC CAUC
1097	UUCCUCAUCCCUGUCC AAAG
1098	UUUAAUGCUCUUUGGG GCUC
1099	UUUCCAGCUUUUUUUC UCCA
1100	AAAGUCCAAGGGGGUU GAAG

SEQ ID NO:	Spacer Sequence
1306	CCCAAAUUCAAGGACA CCUU
1307	CCCAAGAGACGGGGCU GUCC
1308	CCCAGCACUCAGAUCA UGAA
1309	CCCAUCUGUGAAGACG GCAG
1310	CCCUUCUGUGCAGCAG UUUC
1311	CCUGCACUGAGUUUUG GAGA
1312	CCUGGUGUCACAUCAC CUAC
1313	CUACCCGGGAGCAUAA AAUA
1314	CUAUCUAUAAAAUAGG GCUA
1315	CUCAUCUCCAGGAAAC UGAG
1316	CUCCAGUUUUUCAUUU CCCA
1317	CUCCAUUGUACAGCUA UGAA
1318	CUCCCCUUCUAUAGAU GAGG
1319	CUCCUCUCCAGUCGCC CUCC
1320	CUCCUGGCCCUCGUGC CUCU

SEQ ID NO:	Spacer Sequence
1526	UAGCCCUGGCAGGGCC AUCA
1527	UAUAGAUGAGGAGGUU AAGG
1528	UAUCAGAGGCAGCACC CUGC
1529	UAUCUAAUCCCACUAU UUAG
1530	UAUCUAUAAAAUAGGG CUAA
1531	UCAUACAUAAACAUCA CUUU
1532	UCAUCUCCAGGAAACU GAGG
1533	UCCAGUUUUUCAUUUC CCAA
1534	UCCAUCCCCUCCAGAUA UGA
1535	UCCAUCUGUUUUGCUA CAUU
1536	UCCAUUGUACAGCUAU GAAG
1537	UCCCAGCCCCACUUUGC ACA
1538	UCCCCUUCUAUAGAUG AGGA
1539	UCCUCAGCUAGGACCU CAGU
1540	UCCUCUCCAGUCGCCCU CCC

SEQ ID NO:	Spacer Sequence
1746	AGGGGCACUGUCACCCA GGG
1747	UGCAGGUACAGGGAGG GCUU
1748	CCUUGACUCGGGGCCUG GCC
1749	CAGGUACAGGGAGGGCU UCC
1750	GCUGCCGGGCAUAUUCU CCU
1751	GAUGACGUUGACUUUG ACCA
1752	CCAGGCCUCAGCCUUAC AUC
1753	CAGCAGUUUCCCUGGCC ACU
1754	UUGCUGUGGAUGACGU UGAC
1755	CUCCUCCCAGCUGUCAC CCC
1756	GGGACAGCCCCGUCUCU UGG
1757	CAAGAUGCAUUUCCCUU CUG
1758	AGAUUAGGCAGGGAGU GGCU
1759	AUUCCCAUUUUACAGAU GAG
1760	ACUAGUAGUUAACAAU GUGG

SEQ ID NO:	Spacer Sequence
1101	AAAGUUUCCUUCCAAUAAAG
1102	AACCAAAAGAAUGAAGGCUU
1103	AAGCAGCACAAACCCCUGCU
1104	AAGCUUGGCCUCAGUGUCCG
1105	AAGUAGAUGGAGGAGGUGGG
1106	ACUUCAGAUAGACCUGGGAU
1107	AUAAUCCCUAGGCACUUCCA
1108	AUCACAGCCCCUCCAUGCUC
1109	AUCCCAGACAAAACAUUUAG
1110	AUGCUCUUUGGGGCUCAAUU
1111	AUGGAAAAUUUGCCAAACUC
1112	AUUUCACAAUAGCUCUAUGA
1113	AUUUCUCUUUCACUUCCCAA
1114	AUUUUAGCCACAGACUGAUG
1115	AUUUUUCCUGCACUGAGUUU
1321	CUCUUUCACUUCCCAACUGA
1322	CUGAUUCAAGACUUCUGGCU
1323	CUGGGAUCGGGGCAGGUGGG
1324	CUGUUUGCACAGCUCCUCUG
1325	CUUCAGCUAGCUAGUGUUGU
1326	CUUCCCGGAGAGCCUGCUGC
1327	CUUCUUUUGUAAAAACAAUGG
1328	CUUGAACCCGGGAGGUGGAG
1329	CUUGUAACAAUCCCGUGAGG
1330	CUUUCUGUUUGCACAGCUCC
1331	CUUUGUAUUUUGUUAAGAUG
1332	GAAUUUGCAACACCUGCCAA
1333	GACCAUUGCUCGGCAGAGCC
1334	GACCUGUCUCCAAAACUCAG
1335	GACUUCUGUCCCCUCCCCUC
1541	UCCUGGCCCUCGUGCCUCUA
1542	UCGAGAACUUGCUGCGCAGC
1543	UCUCCACCACCUUUCUCCUG
1544	UCUUCCCUGCACCCUGCCGG
1545	UCUUGGGUCUCAGGCGCUGG
1546	UCUUUCACUUCCCAACUGAA
1547	UGACAUUCAUCAUCUGAGAU
1548	UGAGCUCCCCCAUGGCCCAGG
1549	UGAUCUGGCCCGAUUCCUGG
1550	UGAUUCAAGACUUCUGGCUU
1551	UGAUUCUUUCAGUGUUACUG
1552	UGCUUACAGGAACCAGUGUA
1553	UGGCUUCAGCUUUGUACACA
1554	UGGGAUCGGGGCAGGUGGGA
1555	UGGGCUCAAAGGAGGUGGAG
1761	GCAAUGAUGGUGACGCUGAG
1762	AAAGGAGGAAACCAGAGUAG
1763	GGGCAGGUGGGAGCACUCUU
1764	GGCCAGAUCAGAAUCAAGAA
1765	CUCUGCACCAGGCCACGGCU
1766	UGGGCAGGCCCAUUCCUCUU
1767	CUUUAUCUGCCAGGGUGGAG
1768	GCUCACUGAAACCUCCACCU
1769	CCUAUGGUCAAACAACCUGU
1770	AGAAUACCAGUUCUUGCUGC
1771	UGCCUCUAAGUAAUCCAAAU
1772	GGGCCUGGCCCCCCCCAUCUC
1773	CUUCCCUGUCCCUGGCCCCG
1774	AGAACUUGCUGCGCAGCAGG
1775	GGCAGGAGCUCAACCUGUGU

SEQ ID NO:	Spacer Sequence
1116	CAAAAGAAGAAAAUGA GGCA
1117	CAGAUGAAGAAACCAA AGCC
1118	CAGAUGAGGAAGAUGG GGAC
1119	CAGGCACCGUAUCAUC UAAU
1120	CCCCAGGCCCCCUGCCA GUA
1121	CUGCCUGCCAGCUGCC UAAC
1122	GAGACGCUCAUCUCUC CUUG
1123	GAGGCGUUCUAUAGGA AACU
1124	GCAAGAAUCCUCUGGU CCUG
1125	GCACCGUCUUUAGUUC UAAC
1126	GCCACAGACUGAUGCU CCCC
1127	GCCAGGCUGCCUCUGA AGUU
1128	GGAGUGCAGGGUGCUG CCUC
1129	GGAUUACAUCUCAGUU UUGU
1130	GUAGAGCUGAGGUUUC ACCG

SEQ ID NO:	Spacer Sequence
1336	GAGACAGAGUCUCGCU UUAU
1337	GAGCACUAGCUUCAUA GCUG
1338	GAGCCCAGAAUUCCCC UGUC
1339	GAGGACAGAUUAGAUG AUAC
1340	GAGGGAGGGGUUGGCA GGUG
1341	GAUGUAAGGCUGAGGC CUGG
1342	GAUGUAAGGUUUGGUG UUGA
1343	GCAACACCUGCCAACC CCUC
1344	GCACAACUGGGGAGAG UGAG
1345	GCACCUUAUAGGGUAA GCUU
1346	GCACUUCAUUUGUUGC UUUA
1347	GCAUCUUAGGGUCAUU GCAG
1348	GCCAAACUCAGAUGCU GUUC
1349	GCCCAAGAGAACAGAG AGGU
1350	GCCUCAUACAUAAAGU GAUG

SEQ ID NO:	Spacer Sequence
1556	UGGUUCUGCUACUUCC UCAG
1557	UGUCCCCUCCCCUCACC CCA
1558	UGUCUCAGCUUCCCAA AUAG
1559	UGUGCAGCAGUUUCCC UGGC
1560	UGUUUGCACAGCUCCU CUGU
1561	UUCAACCCCCCUUGGAC UUUU
1562	UUCAGCUAGCUAGUGU UGUU
1563	UUCAUCUGUAAAAGAA GGAU
1564	UUCCCGGAGAGCCUGC UGCA
1565	UUCUGAUUCUUUCAGU GUUA
1566	UUCUGUCUCAGCUUCC CAAA
1567	UUCUUUUGUAAAACAA UGGC
1568	UUGAACCCGGGAGGUG GAGG
1569	UUGAUUCUGAUCUGGC CCGA
1570	UUGCUAAAGGCCAAGU GGAU

SEQ ID NO:	Spacer Sequence
1776	GCAGAGCCUGCUCUCGC UGG
1777	CCUGUAACCCCCAGCACU UUG
1778	CUGGUUCAGCCACAGGC CCC
1779	GUGACUUCAGACAGUUG CUC
1780	GUACCUCCUGGUGGAUA CAC
1781	CCACCCUCCCCUGCACCA GCC
1782	GGUGGGGUGAGGGGAG GGGA
1783	CCAUCUGGCCAAGUACC CCA
1784	GCAAGUACUUGGCACAG GCU
1785	GGGGCUCCCUGUCUGUUG UCU
1786	CCCUCCCCAGGCACUAGC UUU
1787	GACCACCAGGCCUGCAGC CCC
1788	GGCUCAGGCACCUGGCU CCA

[682] **TABLE 6** provides exemplary repeat sequences and associated effector proteins.

TABLE 6: Exemplary Repeat Sequences

Effector Protein Seq ID NO:	Repeat Seq ID NO:	Repeat sequence (5' → 3'), shown as RNA
6-12, 228-230	252	AUAGAUUGCUCCUUACGAGGAGAC
13	253	GAAUGAAGGAAUGCAAC
47	254	CCUGAUUGCUCGCUGCGGCGAGAC
231	255	AAGGAUGCCAAAC
181-184	256	UGGGGCAGUUGGUUGCCCUUAGCCUGAGGCAUUUAUUGCACUCGG GAAGUACCAUUUCUCAGAAAUGGUACAUCCAAC
187-188	257	AUGGGGCAGUUGGUUGCCCUUAGCCUGAGGAAUUUAAUUCACUCG GGAAGUACCUUUCUCAUGAAAUGGUACAUCCAAC
181-185	258	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGCACUUGAGUGA AGGUGGGCUGCUUGCAUCAGCCUAAUGUCGAGAAGUGCUUUCUUC GGAAAGUAACCCUCGAAACAAAUUCAUUUGAAAGAAUGAAGGAAU GCAAC
231	1789	GUUUGAGAACCUUAUGAAAUUACAAGGAUGCCAAAC
6-12, 228-230	1848	CUUUCAAGACUAAUAGAUUGCUCCUUACGAGGAGAC

[683] **TABLE 7** provides exemplary handle sequences.

TABLE 7: Exemplary Handle Sequence

Seq ID No:	Handle sequence (5' → 3'), shown as RNA	Description

259	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGc ACUUGAGUGAAGGUGGGCUGCUUGCAUCAGCCUAA UGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCG AAACAAAUUCAUUU	Intermediary sequence
260	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGc ACUUGAGUGAAGGUGGGCUGCUUGCAUCAGCCUAA UGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCG AAACAAAUUCAUUUGAAA	Intermediary sequence + Linker
261	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGc ACUUGAGUGAAGGUGGGCUGCUUGCAUCAGCCUAA UGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCG AAACAAAUUCAUUUGAAAGAAUGAAGGAAUGCAAC	Intermediary sequence + Linker + Repeat Sequence
262	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUA ACACUCACAAGAAUCCU	Intermediary sequence
263	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUA ACACUCACAAGAAUCCUGAAA	Intermediary sequence + Linker
264	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUA ACACUCACAAGAAUCCUGAAAAAGGAUGCCAAAC	Intermediary sequence + Linker + Repeat Sequence

[684] **TABLE 8** provides exemplary linker sequences.

TABLE 8: Exemplary Linker Sequences

SEQ ID NO:	Linker sequence (5' → 3'), shown as RNA
265	GAAA

[685] **TABLE 9** provides exemplary guide sequences and associated effector proteins.

TABLE 9. Exemplary Guide RNA Sequences

Effector Protein SEQ ID NO:	SEQ ID NO:	Type of gRNA	gRNA sequence
6-12, 228-230	605	crRNA	AUAGAUUGCUCCUUACGAGGAGACAACAAUAGAAAAAUGGA
6-12, 228-230	606	crRNA	AUAGAUUGCUCCUUACGAGGAGACCUGGGAAAAUAGAAUAA
6-12, 228-230	607	crRNA	AUAGAUUGCUCCUUACGAGGAGACACUGGGAAAAUAGAAUA
6-12, 228-230	608	crRNA	AUAGAUUGCUCCUUACGAGGAGACUACUGGGAAAAUAGAAU
6-12, 228-230	609	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUCUAUUGUUCAACUUU
6-12, 228-230	610	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUUUACUGGGAAAAUAG
6-12, 228-230	611	crRNA	AUAGAUUGCUCCUUACGAGGAGACUCUAUUGUUCAACUUUU
6-12, 228-230	612	crRNA	AUAGAUUGCUCCUUACGAGGAGACAUUUUACUGGGAAAAUA
6-12, 228-230	613	crRNA	AUAGAUUGCUCCUUACGAGGAGACCUAUUGUUCAACUUUUA
6-12, 228-230	614	crRNA	AUAGAUUGCUCCUUACGAGGAGACUAUUGUUCAACUUUUAU
6-12, 228-230	615	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUCAACUUUUAUUCUAU
6-12, 228-230	616	crRNA	AUAGAUUGCUCCUUACGAGGAGACAACUUUUAUUCUAUUUU
6-12, 228-230	617	crRNA	AUAGAUUGCUCCUUACGAGGAGACCUAAAACUUUAUUUUAC
6-12, 228-230	618	crRNA	AUAGAUUGCUCCUUACGAGGAGACACUAAAACUUUAUUUUA
6-12, 228-230	619	crRNA	AUAGAUUGCUCCUUACGAGGAGACUAUUCUAUUUUCCCAGU
6-12, 228-230	620	crRNA	AUAGAUUGCUCCUUACGAGGAGACAUUCUAUUUUCCCAGUA
6-12, 228-230	621	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUCUAUUUUCCCAGUAA
6-12, 228-230	622	crRNA	AUAGAUUGCUCCUUACGAGGAGACUAUUUUCCCAGUAAAAU
6-12, 228-230	623	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAGAUGCAGAGUUUACU
6-12, 228-230	624	crRNA	AUAGAUUGCUCCUUACGAGGAGACUCCCAGUAAAAUAAAGU
6-12, 228-230	625	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAAGAUGCAGAGUUUAC
6-12, 228-230	626	crRNA	AUAGAUUGCUCCUUACGAGGAGACCCCAGUAAAAUAAAGUU
6-12, 228-230	627	crRNA	AUAGAUUGCUCCUUACGAGGAGACCCAGUAAAAUAAAGUUU
6-12, 228-230	628	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUUUAAAGAUGCAGAGUU
6-12, 228-230	629	crRNA	AUAGAUUGCUCCUUACGAGGAGACUAGUAAAACUCUGCAUCU

Effector Protein SEQ ID NO:	SEQ ID NO:	Type of gRNA	gRNA sequence
6-12, 228-230	630	crRNA	AUAGAUUGCUCCUUACGAGGAGACAGUAAACUCUGCAUCUU
6-12, 228-230	631	crRNA	AUAGAUUGCUCCUUACGAGGAGACGUAAACUCUGCAUCUUU
6-12, 228-230	632	crRNA	AUAGAUUGCUCCUUACGAGGAGACGAAAUAAAUGCCAAAAU
6-12, 228-230	633	crRNA	AUAGAUUGCUCCUUACGAGGAGACAGAAAUAAAUGCCAAAA
6-12, 228-230	634	crRNA	AUAGAUUGCUCCUUACGAGGAGACUAGAAAUAAAUGCCAAA
6-12, 228-230	635	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAAGAAUUAUUUUGGCA
6-12, 228-230	636	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAGAAUUAUUUUGGCAU
6-12, 228-230	637	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUUUGGCAUUUAUUUCU
6-12, 228-230	638	crRNA	AUAGAUUGCUCCUUACGAGGAGACUGGCAUUUAUUUCUAAA
6-12, 228-230	639	crRNA	AUAGAUUGCUCCUUACGAGGAGACGGCAUUUAUUUCUAAAA
6-12, 228-230	640	crRNA	AUAGAUUGCUCCUUACGAGGAGACGCAUUUAUUUCUAAAAU
6-12, 228-230	641	crRNA	AUAGAUUGCUCCUUACGAGGAGACACAAAUACAAAAUACUA
6-12, 228-230	642	crRNA	AUAGAUUGCUCCUUACGAGGAGACAUUUCUAAAAAUGGCAUA
6-12, 228-230	643	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUUCUAAAAUGGCAUAG
6-12, 228-230	644	crRNA	AUAGAUUGCUCCUUACGAGGAGACUAAGACUUCACAAAUAC
6-12, 228-230	645	crRNA	AUAGAUUGCUCCUUACGAGGAGACCUAAAAUGGCAUAGUAU
6-12, 228-230	646	crRNA	AUAGAUUGCUCCUUACGAGGAGACUAAAAUGGCAUAGUAUU
6-12, 228-230	647	crRNA	AUAGAUUGCUCCUUACGAGGAGACAUAAGAUAACCUUGUAA
6-12, 228-230	648	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUAAUAAGAUAACCUUG
6-12, 228-230	649	crRNA	AUAGAUUGCUCCUUACGAGGAGACAUUAAUAAGAUAACCUU
6-12, 228-230	650	crRNA	AUAGAUUGCUCCUUACGAGGAGACUAUUAAUAAGAUAACCU
6-12, 228-230	651	crRNA	AUAGAUUGCUCCUUACGAGGAGACUGUAUUUGUGAAGUCUU
6-12, 228-230	652	crRNA	AUAGAUUGCUCCUUACGAGGAGACGUAUUUGUGAAGUCUUA
6-12, 228-230	653	crRNA	AUAGAUUGCUCCUUACGAGGAGACUAUUUGUGAAGUCUUAC
6-12, 228-230	654	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAUUUUAUUAAUAAGAU

Effector Protein SEQ ID NO:	SEQ ID NO:	Type of gRNA	gRNA sequence
6-12, 228-230	655	crRNA	AUAGAUUGCUCCUUACGAGGAGACGAAUUUUAUUAAUAAGA
6-12, 228-230	656	crRNA	AUAGAUUGCUCCUUACGAGGAGACGUGAAGUCUUACAAGGU
6-12, 228-230	657	crRNA	AUAGAUUGCUCCUUACGAGGAGACUGAAGUCUUACAAGGUU
6-12, 228-230	658	crRNA	AUAGAUUGCUCCUUACGAGGAGACCCUAGGAUGUUUGAAUU
6-12, 228-230	659	crRNA	AUAGAUUGCUCCUUACGAGGAGACCAAGGUUAUCUUAUUAA
6-12, 228-230	660	crRNA	AUAGAUUGCUCCUUACGAGGAGACACCUAGGAUGUUUGAAU
6-12, 228-230	661	crRNA	AUAGAUUGCUCCUUACGAGGAGACUACCUAGGAUGUUUGAA
6-12, 228-230	662	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUACCUAGGAUGUUUGA
6-12, 228-230	663	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUUACCUAGGAUGUUUG
6-12, 228-230	664	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUUUACCUAGGAUGUUU
6-12, 228-230	665	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUUUUACCUAGGAUGUU
6-12, 228-230	666	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUUUUUACCUAGGAUGU
6-12, 228-230	667	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUUUUUUACCUAGGAUG
6-12, 228-230	668	crRNA	AUAGAUUGCUCCUUACGAGGAGACUCUUAUUAAUAAAAUUC
6-12, 228-230	669	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUUUUUUUUACCUAGGAU
6-12, 228-230	670	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUUUUUUUUUACCUAGGA
6-12, 228-230	671	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUUUUUUUUUUACCUAGG
6-12, 228-230	672	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUAAUAAAAUUCAAACA
6-12, 228-230	673	crRNA	AUAGAUUGCUCCUUACGAGGAGACAUAAAAAUUCAAACAUCC
6-12, 228-230	674	crRNA	AUAGAUUGCUCCUUACGAGGAGACUGACCUUUUUUUUUUUUU
6-12, 228-230	675	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAACAUCCUAGGUAAAA
6-12, 228-230	676	crRNA	AUAGAUUGCUCCUUACGAGGAGACCAGUCACUAAACAAUUC
6-12, 228-230	677	crRNA	AUAGAUUGCUCCUUACGAGGAGACGUGCGCAAAAGAAAAUU
6-12, 228-230	678	crRNA	AUAGAUUGCUCCUUACGAGGAGACCUUAGUGCGCAAAAGAA
6-12, 228-230	679	crRNA	AUAGAUUGCUCCUUACGAGGAGACCCUUAGUGCGCAAAAGA

Effector Protein SEQ ID NO:	SEQ ID NO:	Type of gRNA	gRNA sequence
6-12, 228-230	680	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUUAGUGACUGUAAUUU
6-12, 228-230	681	crRNA	AUAGAUUGCUCCUUACGAGGAGACCACUUUCCUUAGUGCGC
6-12, 228-230	682	crRNA	AUAGAUUGCUCCUUACGAGGAGACAGUGACUGUAAUUUUCU
6-12, 228-230	683	crRNA	AUAGAUUGCUCCUUACGAGGAGACGCACUUUCCUUAGUGCG
6-12, 228-230	684	crRNA	AUAGAUUGCUCCUUACGAGGAGACGUGACUGUAAUUUUCUU
6-12, 228-230	685	crRNA	AUAGAUUGCUCCUUACGAGGAGACCUUUGCACUUUCCUUAG
6-12, 228-230	686	crRNA	AUAGAUUGCUCCUUACGAGGAGACUCUUUUGCGCACUAAGG
6-12, 228-230	687	crRNA	AUAGAUUGCUCCUUACGAGGAGACCUUUUGCGCACUAAGGA
6-12, 228-230	688	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUUUGCGCACUAAGGAA
6-12, 228-230	689	crRNA	AUAGAUUGCUCCUUACGAGGAGACUGCGCACUAAGGAAAGU
6-12, 228-230	690	crRNA	AUAGAUUGCUCCUUACGAGGAGACAGUCACUCUAAGUUACU
6-12, 228-230	691	crRNA	AUAGAUUGCUCCUUACGAGGAGACGCGCACUAAGGAAAGUG
6-12, 228-230	692	crRNA	AUAGAUUGCUCCUUACGAGGAGACCAGUCACUCUAAGUUAC
6-12, 228-230	693	crRNA	AUAGAUUGCUCCUUACGAGGAGACCGCACUAAGGAAAGUGC
6-12, 228-230	694	crRNA	AUAGAUUGCUCCUUACGAGGAGACUGUGAAGUUUCAGUCAC
6-12, 228-230	695	crRNA	AUAGAUUGCUCCUUACGAGGAGACAACCCUAUUCUGUGAAG
6-12, 228-230	696	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAUCUUCAACCCUAUUC
6-12, 228-230	697	crRNA	AUAGAUUGCUCCUUACGAGGAGACGAGUGACUGAAACUUCA
6-12, 228-230	698	crRNA	AUAGAUUGCUCCUUACGAGGAGACUGAAUUCAAUCUUCAAC
6-12, 228-230	699	crRNA	AUAGAUUGCUCCUUACGAGGAGACGGAUAGUUAUGAAUUCA
6-12, 228-230	700	crRNA	AUAGAUUGCUCCUUACGAGGAGACGGGAUAGUUAUGAAUUC
6-12, 228-230	701	crRNA	AUAGAUUGCUCCUUACGAGGAGACACAGAAUAGGGUUGAAG
6-12, 228-230	702	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAGAUUGAAUUCAUAAC
6-12, 228-230	703	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAUUCAUAACUAUCCCA
6-12, 228-230	704	crRNA	AUAGAUUGCUCCUUACGAGGAGACAUAACUAUCCCAAAGAC

Effector Protein SEQ ID NO:	SEQ ID NO:	Type of gRNA	gRNA sequence
6-12, 228-230	705	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAUAAAGCAUAGUGCAA
6-12, 228-230	706	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAAUAAAGCAUAGUGCA
6-12, 228-230	707	crRNA	AUAGAUUGCUCCUUACGAGGAGACUAAAUAAAGCAUAGUGC
6-12, 228-230	708	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUAAAUAAAGCAUAGUG
6-12, 228-230	709	crRNA	AUAGAUUGCUCCUUACGAGGAGACUGGUUUUUAAAUAAAGC
6-12, 228-230	710	crRNA	AUAGAUUGCUCCUUACGAGGAGACGUGGUUUUUAAAUAAAG
6-12, 228-230	711	crRNA	AUAGAUUGCUCCUUACGAGGAGACUGUGGUUUUUAAAUAAA
6-12, 228-230	712	crRNA	AUAGAUUGCUCCUUACGAGGAGACCACUAUGCUUUAUUUAA
6-12, 228-230	713	crRNA	AUAGAUUGCUCCUUACGAGGAGACUGAGAUCAACAGCACAG
6-12, 228-230	714	crRNA	AUAGAUUGCUCCUUACGAGGAGACAUGAGAUCAACAGCACA
6-12, 228-230	715	crRNA	AUAGAUUGCUCCUUACGAGGAGACAUUUAAAAACCACAAAA
6-12, 228-230	716	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUUAAAAACCACAAAAC
6-12, 228-230	717	crRNA	AUAGAUUGCUCCUUACGAGGAGACUAUUUAUGAGAUCAACA
6-12, 228-230	718	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAAAACCACAAAACCUG
6-12, 228-230	719	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAAACCACAAAACCUGU
6-12, 228-230	720	crRNA	AUAGAUUGCUCCUUACGAGGAGACAUCUCAUAAAAUAGAACU
6-12, 228-230	721	crRNA	AUAGAUUGCUCCUUACGAGGAGACUAUUUAUAUUUAUUUUC
6-12, 228-230	722	crRNA	AUAGAUUGCUCCUUACGAGGAGACAUAUUUAUUUUCAUUUU
6-12, 228-230	723	crRNA	AUAGAUUGCUCCUUACGAGGAGACUAUUUAUUUUCAUUUUA
6-12, 228-230	724	crRNA	AUAGAUUGCUCCUUACGAGGAGACAUUUUCAUUUUAGUCUG
6-12, 228-230	725	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUUUUCAUUUUAGUCUGU
6-12, 228-230	726	crRNA	AUAGAUUGCUCCUUACGAGGAGACUCAUUUUAGUCUGUCUU
6-12, 228-230	727	crRNA	AUAGAUUGCUCCUUACGAGGAGACCAUUUUAGUCUGUCUUC
6-12, 228-230	728	crRNA	AUAGAUUGCUCCUUACGAGGAGACGUGUCUAUCAACAGCAA
6-12, 228-230	729	crRNA	AUAGAUUGCUCCUUACGAGGAGACAUUUUAGUCUGUCUUCU

Effector Protein SEQ ID NO:	SEQ ID NO:	Type of gRNA	gRNA sequence
6-12, 228-230	730	crRNA	AUAGAUUGCUCCUUACGAGGAGACAGUGUCUAUCAACAGCA
6-12, 228-230	731	crRNA	AUAGAUUGCUCCUUACGAGGAGACUAGUGUCUAUCAACAGC
6-12, 228-230	732	crRNA	AUAGAUUGCUCCUUACGAGGAGACUAGUCUGUCUUCUUGGU
6-12, 228-230	733	crRNA	AUAGAUUGCUCCUUACGAGGAGACAGUCUGUCUUCUUGGUU
6-12, 228-230	734	crRNA	AUAGAUUGCUCCUUACGAGGAGACGUCUGUCUUCUUGGUUG
6-12, 228-230	735	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUGGUUGCUGUUGAUAG
6-12, 228-230	736	crRNA	AUAGAUUGCUCCUUACGAGGAGACGUUGCUGUUGAUAGACA
6-12, 228-230	737	crRNA	AUAGAUUGCUCCUUACGAGGAGACGAUAAUAUCUAAUACUC
6-12, 228-230	738	crRNA	AUAGAUUGCUCCUUACGAGGAGACCUGUUGAUAGACACUAA
6-12, 228-230	739	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAACUUAGAUAAAUAUCU
6-12, 228-230	740	crRNA	AUAGAUUGCUCCUUACGAGGAGACAUAGACACUAAAAGAGU
6-12, 228-230	741	crRNA	AUAGAUUGCUCCUUACGAGGAGACUAUUCAAACUUAGAUAA
6-12, 228-230	742	crRNA	AUAGAUUGCUCCUUACGAGGAGACUAGCCUUAUAUUCAAAC
6-12, 228-230	743	crRNA	AUAGAUUGCUCCUUACGAGGAGACAUAGCCUUAUAUUCAAA
6-12, 228-230	744	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAUAUUUAUAGCCUUAU
6-12, 228-230	745	crRNA	AUAGAUUGCUCCUUACGAGGAGACGAUAUUAUCUAAGUUUG
6-12, 228-230	746	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUAAAUAUUUAUAGCCU
6-12, 228-230	747	crRNA	AUAGAUUGCUCCUUACGAGGAGACUCUAAGUUUGAAUAUAA
6-12, 228-230	748	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAAAUUAUUAAAUAUUU
6-12, 228-230	749	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAAAUUAUUAAAUAUU
6-12, 228-230	750	crRNA	AUAGAUUGCUCCUUACGAGGAGACUAAAAAUUAUUAAAUAU
6-12, 228-230	751	crRNA	AUAGAUUGCUCCUUACGAGGAGACGAAUAUAAGGCUAUAAA
6-12, 228-230	752	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAUAUAAGGCUAUAAAU
6-12, 228-230	753	crRNA	AUAGAUUGCUCCUUACGAGGAGACCCAAGAAUACUAUUUUA
6-12, 228-230	754	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAUUACCAAGAAUACUA

Effector Protein SEQ ID NO:	SEQ ID NO:	Type of gRNA	gRNA sequence
6-12, 228-230	755	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAUAAUUUUUAAAAUAG
6-12, 228-230	756	crRNA	AUAGAUUGCUCCUUACGAGGAGACAUAAUUUUUAAAAUAGU
6-12, 228-230	757	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUAAAAUAGUAUUCUUG
6-12, 228-230	758	crRNA	AUAGAUUGCUCCUUACGAGGAGACUAAAAUAGUAUUCUUGG
6-12, 228-230	759	crRNA	AUAGAUUGCUCCUUACGAGGAGACAACAGAAGAAUAAUUCA
6-12, 228-230	760	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAAAUAGUAUUCUUGGU
6-12, 228-230	761	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAACAGAAGAAUAAUUC
6-12, 228-230	762	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAAUAGUAUUCUUGGUA
6-12, 228-230	763	crRNA	AUAGAUUGCUCCUUACGAGGAGACUGCCUUUAAACAGAAGA
6-12, 228-230	764	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUCUGCCUUUAAACAGA
6-12, 228-230	765	crRNA	AUAGAUUGCUCCUUACGAGGAGACCUUCUGCCUUUAAACAG
6-12, 228-230	766	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUGGUAAUUGAAUUAUU
6-12, 228-230	767	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUUCUUCUGCCUUUAAA
6-12, 228-230	768	crRNA	AUAGAUUGCUCCUUACGAGGAGACGUAAUUGAAUUAUUCUU
6-12, 228-230	769	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAUUAUUUCUUCUGCCU
6-12, 228-230	770	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAUUAUUCUUCUGUUUA
6-12, 228-230	771	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUCUUCUGUUUAAAGGC
6-12, 228-230	772	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUCUGUUUAAAGGCAGA
6-12, 228-230	773	crRNA	AUAGAUUGCUCCUUACGAGGAGACUGUUUAAAGGCAGAAGA
6-12, 228-230	774	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAAGGCAGAAGAAAUAA
6-12, 228-230	775	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAGGCAGAAGAAAUAAU
6-12, 228-230	776	crRNA	AUAGAUUGCUCCUUACGAGGAGACCUACAGAAAAACUCAGG
6-12, 228-230	777	crRNA	AUAGAUUGCUCCUUACGAGGAGACGGCUCUGAUUCCUACAG
6-12, 228-230	778	crRNA	AUAGAUUGCUCCUUACGAGGAGACAACAUCAUCCUGAGUUU
6-12, 228-230	779	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAAAUAUUGGGCUCUGA

Effector Protein SEQ ID NO:	SEQ ID NO:	Type of gRNA	gRNA sequence
6-12, 228-230	780	crRNA	AUAGAUUGCUCCUUACGAGGAGACCAAAAUAUUGGGCUCUG
6-12, 228-230	781	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUUCAAAAUAUUGGGCU
6-12, 228-230	782	crRNA	AUAGAUUGCUCCUUACGAGGAGACGUUUCAAAAUAUUGGGC
6-12, 228-230	783	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUCUGUAGGAAUCAGAG
6-12, 228-230	784	crRNA	AUAGAUUGCUCCUUACGAGGAGACUCUGUAGGAAUCAGAGC
6-12, 228-230	785	crRNA	AUAGAUUGCUCCUUACGAGGAGACCUGUAGGAAUCAGAGCC
6-12, 228-230	786	crRNA	AUAGAUUGCUCCUUACGAGGAGACUGCAUUUGUUUCAAAAU
6-12, 228-230	787	crRNA	AUAGAUUGCUCCUUACGAGGAGACUGUAGGAAUCAGAGCCC
6-12, 228-230	788	crRNA	AUAGAUUGCUCCUUACGAGGAGACGAUUAUGCAUUUGUUUC
6-12, 228-230	789	crRNA	AUAGAUUGCUCCUUACGAGGAGACACUUAGAUUAUGCAUUU
6-12, 228-230	790	crRNA	AUAGAUUGCUCCUUACGAGGAGACGACUUAGAUUAUGCAUU
6-12, 228-230	791	crRNA	AUAGAUUGCUCCUUACGAGGAGACCAUUUGACUUAGAUUAU
6-12, 228-230	792	crRNA	AUAGAUUGCUCCUUACGAGGAGACCCAUUUGACUUAGAUUA
6-12, 228-230	793	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUUCCAUUUGACUUAGA
6-12, 228-230	794	crRNA	AUAGAUUGCUCCUUACGAGGAGACCUUUCCAUUUGACUUAG
6-12, 228-230	795	crRNA	AUAGAUUGCUCCUUACGAGGAGACUGAAACAAAUGCAUAAU
6-12, 228-230	796	crRNA	AUAGAUUGCUCCUUACGAGGAGACGAAACAAAUGCAUAAUC
6-12, 228-230	797	crRNA	AUAGAUUGCUCCUUACGAGGAGACUAUUUCUUUCCAUUUGA
6-12, 228-230	798	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAACAAAUGCAUAAAUCU
6-12, 228-230	799	crRNA	AUAGAUUGCUCCUUACGAGGAGACAUAUUUCUUUCCAUUUG
6-12, 228-230	800	crRNA	AUAGAUUGCUCCUUACGAGGAGACUAUAUUUCUUUCCAUUU
6-12, 228-230	801	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUAUAUUUCUUUCCAUU
6-12, 228-230	802	crRNA	AUAGAUUGCUCCUUACGAGGAGACCUUUUUAUAUUUCUUUC
6-12, 228-230	803	crRNA	AUAGAUUGCUCCUUACGAGGAGACAAUACUGAAGAAAACAA
6-12, 228-230	804	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUAAAAUACUGAAGAAAA

Effector Protein SEQ ID NO:	SEQ ID NO:	Type of gRNA	gRNA sequence
6-12, 228-230	805	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUACUUCUUGUUUUCUU
6-12, 228-230	806	crRNA	AUAGAUUGCUCCUUACGAGGAGACCUUCUUGUUUUCUUCAG
6-12, 228-230	807	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUGUUUUCUUCAGUAUU
6-12, 228-230	808	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUUUCUUCAGUAUUUAA
6-12, 228-230	809	crRNA	AUAGAUUGCUCCUUACGAGGAGACUCUUCAGUAUUUAACAA
6-12, 228-230	810	crRNA	AUAGAUUGCUCCUUACGAGGAGACCUUCAGUAUUUAACAAU
6-12, 228-230	811	crRNA	AUAGAUUGCUCCUUACGAGGAGACUUCAGUAUUUAACAAUC
6-12, 228-230	812	crRNA	AUAGAUUGCUCCUUACGAGGAGACAGUAUUUAACAAUCCUU
6-12, 228-230	813	crRNA	AUAGAUUGCUCCUUACGAGGAGACAACAAUCCUUUUUUUUC
6-12, 228-230	814	crRNA	AUAGAUUGCUCCUUACGAGGAGACACAAUCCUUUUUUUUCU
6-12, 228-230	1847	crRNA	CUUUCAAGACUAAUAGAUUGCUCCUUACGAGGAGACAGUAUUUAAUUAAAUACUCGUAUUGCU
47	815	crRNA	CCUGAUUGCUCGGCUGCGGCGAGACAACAAUAGAAAAAUGGAUUU
47	816	crRNA	CCUGAUUGCUCGCUGCGGCGAGACAACUUUUAUUCUAUUUUUCCC
47	817	crRNA	CCUGAUUGCUCGCUGCGGCGAGACACUAAAACUUUAUUUUUACUG
47	818	crRNA	CCUGAUUGCUCGCUGCGGCGAGACUAGUAAAACUCUGCAUCUUUA
47	819	crRNA	CCUGAUUGCUCGCUGCGGCGAGACGAAUUUUAUUAAUUAAGAUAA
47	820	crRNA	CCUGAUUGCUCGCUGCGGCGAGACUCUUAUUAAUAAAAUUUCAAA
47	821	crRNA	CCUGAUUGCUCGCUGCGGCGAGACAGUGACUGUAAUUUUCUUUU
47	822	crRNA	CCUGAUUGCUCGCUGCGGCGAGACCUUUGCACUUUCCUUAGUGC
47	823	crRNA	CCUGAUUGCUCGCUGCGGCGAGACCAGUCACUCUAAGUUACUUU
47	824	crRNA	CCUGAUUGCUCGCUGCGGCGAGACUGAAUUCAAUCUUCAACCCU
47	825	crRNA	CCUGAUUGCUCGCUGCGGCGAGACAAGAUUGAAUUCAUAACUAU
47	826	crRNA	CCUGAUUGCUCGCUGCGGCGAGACUUAAAUAAAGCAUAGUGCAA
47	827	crRNA	CCUGAUUGCUCGCUGCGGCGAGACUGUGGUUUUUAAAUAAAGCA
47	828	crRNA	CCUGAUUGCUCGCUGCGGCGAGACUAUUUAUGAGAUCAACAGCA

Effector Protein SEQ ID NO:	SEQ ID NO:	Type of gRNA	gRNA sequence
47	829	crRNA	CCUGAUUGCUCGCUGCGGCGAGACAUCUCAUAAAUAGAACUUGU
47	830	crRNA	CCUGAUUGCUCGCUGCGGCGAGACCUGUUGAUAGACACUAAAAAG
47	831	crRNA	CCUGAUUGCUCGCUGCGGCGAGACAUAGACACUAAAAGAGUAUU
47	832	crRNA	CCUGAUUGCUCGCUGCGGCGAGACGAAUAUAAGGCUAUAAAUAU
47	833	crRNA	CCUGAUUGCUCGCUGCGGCGAGACAAUUAUUUCUUCUGCCUUUA
47	834	crRNA	CCUGAUUGCUCGCUGCGGCGAGACAAAGGCAGAAGAAAUAAUUG
47	835	crRNA	CCUGAUUGCUCGCUGCGGCGAGACCAAAAUAUUGGGCUCUGAUU
47	836	crRNA	CCUGAUUGCUCGCUGCGGCGAGACUUCUGUAGGAAUCAGAGCCC
47	837	crRNA	CCUGAUUGCUCGCUGCGGCGAGACCUUUUUUAUAUUUCUUUCCAU
47	838	crRNA	CCUGAUUGCUCGCUGCGGCGAGACAAUACUGAAGAAAACAAGAA
47	839	crRNA	CCUGAUUGCUCGCUGCGGCGAGACUCUUCAGUAUUUAACAAUCC
231	840	sgRNA	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAGGA UGCCAAACCAUCUUUAAAGAAUUAUUUU
231	841	sgRNA	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAGGA UGCCAAACCCAUUUUAGAAAUAAAUGCC
231	842	sgRNA	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAGGA UGCCAAACGCAUUUAUUUCUAAAAUGGC
231	843	sgRNA	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAGGA UGCCAAACUAUUUGUGAAGUCUUACAAG
231	844	sgRNA	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAGGA UGCCAAACAAUUUUAUUAAUAAGAUAAC
231	845	sgRNA	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAGGA UGCCAAACUGAAGUCUUACAAGGUUAUC
231	846	sgRNA	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAGGA UGCCAAACAAGUCUUACAAGGUUAUCUU
231	847	sgRNA	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAGGA UGCCAAACACCUUUUUUUUUUUUUACCU
231	848	sgRNA	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAGGA UGCCAAACCACUUUCCUUAGUGCGCAAA

Effector Protein SEQ ID NO:	SEQ ID NO:	Type of gRNA	gRNA sequence
231	849	sgRNA	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAGGA UGCCAAACCGCACUAAGGAAAGUGCAAA
231	850	sgRNA	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAGGA UGCCAAACAAGUUUCAGUCACUCUAAGU
231	851	sgRNA	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAGGA UGCCAAACUGAAGUUUUCAGUCACUCUAA
231	852	sgRNA	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAGGA UGCCAAACAAUUCAAUCUUCAACCCUAU
231	853	sgRNA	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAGGA UGCCAAACGGAUAGUUAUGAAUUCAAUC
231	854	sgRNA	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAGGA UGCCAAACGUUUUUAAAUAAAGCAUAGU
231	855	sgRNA	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAGGA UGCCAAACUGGUUUUUAAAUAAAGCAUA
231	856	sgRNA	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAGGA UGCCAAACAGAUCAACACAGCACAGGUUUU
231	857	sgRNA	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAGGA UGCCAAACCUUUAUUUAAAAACCACAAA
231	858	sgRNA	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAGGA UGCCAAACCUGUUGAUCUCAUAAAUAGA
231	859	sgRNA	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAGGA UGCCAAACUCUUCUUGGUUGCUGUUGAU
231	860	sgRNA	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAGGA UGCCAAACAAUAUAAGGCUAUAAAAUAUU
231	861	sgRNA	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAGGA UGCCAAACCCUUUAAACAGAAGAAUAAU
231	862	sgRNA	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAGGA UGCCAAACUUUAAAGGCAGAAGAAAUAA
231	863	sgRNA	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAGGA UGCCAAACAUUCCUACAGAAAAACUCAG
231	864	sgRNA	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAGGA UGCCAAACUUUCAAAAAUAUUGGGCUCUG

Effector Protein SEQ ID NO:	SEQ ID NO:	Type of gRNA	gRNA sequence
231	865	sgRNA	ACAGCUUAUUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAAGGA UGCCAAACCAUUUUGUUUCAAAAUAUUGG
231	866	sgRNA	ACAGCUUAUUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAAGGA UGCCAAACUAGGAAUCAGAGCCCAAUAU
231	867	sgRNA	ACAGCUUAUUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAAGGA UGCCAAACACUUAGAUUAUGCAUUUUGUU
231	868	sgRNA	ACAGCUUAUUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCACAAGAAUCCUGAAAAAGGA UGCCAAACAAACAAAUGCAUAAUCUAAG
13	869	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACCUGGGAAAAUAGAAUAAAAG
13	870	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACUUUUACUGGGAAAAUAGAAU
13	871	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACCUAAAAACUUUAUUUUUACUGG
13	872	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACUUCUAUUUUUCCCCAGUAAAAU
13	873	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACAAGAUGCAGAGUUUACUAAA
13	874	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACGUAAACUCUGCAUCUUUAAA
13	875	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACGAAAUAAAUGCCAAAAUAAU
13	876	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACAAGAAUUAUUUUGGCAUUUA

Effector Protein SEQ ID NO:	SEQ ID NO:	Type of gRNA	gRNA sequence
13	877	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACGCAUUUAUUUCUAAAAUGGC
13	878	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACUUUCUAAAAAUGGCAUAGUAU
13	879	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACUUAAUAAGAUAACCUUGUAA
13	880	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACUAUUUGUGAAGUCUUACAAG
13	881	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACAAUUUUAUUAAUAAGAUAAC
13	882	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACUGAAGUCUUACAAGGUUAUC
13	883	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACCCUAGGAUGUUUGAAUUUUA
13	884	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACCACACUUUCCUUAGUGCGCAAA
13	885	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACGUGAAUUUUUCUUUUG
13	886	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACCGCACUAAGGAAAGUGCAAA

Effector Protein SEQ ID NO:	SEQ ID NO:	Type of gRNA	gRNA sequence
13	887	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACGGAUAGUUAUGAAUUCAAUC
13	888	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACAAUAAAGCAUAGUGCAAUGG
13	889	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACUGGUUUUUAAAUAAAGCAUA
13	890	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACUGAGAUCAACAGCACAGGUU
13	891	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACUUUAAAAACCACACAAAACCUG
13	892	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACAAAACCACAAAACCUGUGCU
13	893	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACUAUUUAUUUUCAUUUUAGUC
13	894	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACUUUUCAUUUUAGUCUGUCUU
13	895	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACGUGUCUAUCAACAGCAACCA
13	896	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACGUCUGUCUUCUUGGUUGCUG

Effector Protein SEQ ID NO:	SEQ ID NO:	Type of gRNA	gRNA sequence
13	897	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACUAGCCUUAUAUUCAAACUUA
13	898	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACAAAAUUAUUUAAUA
13	899	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACAAUAUAAGGCUAUAAAUAUU
13	900	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACAUAAUUUUUAAAAUAGUAUU
13	901	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACAGAAGAAUAAUUCAAUU
13	902	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACAAAUAGUAUUCUUGGUAAUU
13	903	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACAAGGCAGAAGAAAUAAUUGA
13	904	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACUUUCAAAAUAUUGGGCUCUG
13	905	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACACUUAGAUUAAUGCAUUUUGUU
13	906	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACUAUUUCUUUCCAUUUGACUU

Effector Protein SEQ ID NO:	SEQ ID NO:	Type of gRNA	gRNA sequence
13	907	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACAAACAAAUGCAUAAUCUAAG
13	908	sgRNA	ACCGCUUCACCAAGUGCUGUCCCUUAGGGGAUUAGcACUUGAGUGAAGGUGGGCUGCUUGCA UCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUGAA AGAAUGAAGGAAUGCAACACAAUCCUUUUUUUUCUUCC

[686] **TABLE 10** provides exemplary safe harbor nucleotide sequences.

TABLE 10. Exemplary Safe Harbor Sequence

Safe Harbor Sequences & Description
Intron 1 *ALB*
gtaagaaatccatttttctattgttcaacttttattctattttcccagtaaaataaagttttagtaaactctgcatctttaaagaattattttggcatttatttctaaaatggcatagtattttgtatttgtgaagtcttacaaggttatcttattaataaaatt caaacatcctaggtaaaaaaaaaaaaaggtcagaattgtttagtgactgtaattttctttgcgcactaaggaaagtgcaaagtaacttagagtgactgaaacttcacagaataggg ttgaagattgaattcataactatcccaaagacctatccattgcactatgctttatttaaaaaccacaaaacctgtgctgttgatctcataaatagaacttgtatttatatttatttt catttttagtctgtcttcttggttgctgttgatagacactaaaagagtattagatattatctaagtttgaatataaggctataaatatttaataatttttaaaatagtattcttggtaattgaattattcttctgtttaaaggcagaagaaataattgaacatcatcctgagtttttctgtaggaatcagagcccaatattttgaaacaaatgcataatctaagtcaaatggaaagaaatataaaaagtaacattattacttcttgttttcttcagtatttaacaatccttttttttcttcccttgcccag (**SEQ ID NO: 266**)
CCR5
CTTCAGATAGATTATATCTGGAGTGAAGAATCCTGCCACCTATGTATCTGGCATAGTATTCTGTGTAGTGGGATGAGCAGAGAACAAAAAC AAAATAATCCAGTGAGAAAAGCCCGTAAATAAACCTTCAGACCAGAGATCTATTCTCTAGCTTATTTTAAGCTCAACTTAAAAAGAAGAAC TGTTCTCTGATTCTTTTCGCCTTCAATACACTTAATGATTTAACTCCACCCTCCTTCAAAAGAAACAGCATTTCCTACTTTTATACTGTCTATA TGATTGATTTGCACAGCTCATCTGGCCAGAAGAGCTGAGACATCCGTTCCCCTACAAGAAACTCTCCCCGGGTGGAACAAGATGGATTATC AAGTGTCAAGTCCAATCTATGACATCAATTATTATACATCGGAGCCCTGCCAAAAAATCAATGTGAAGCAAATCGCAGCCCGCCTCCTGCC TCCGCTCTACTCACTGGTGTTCATCTTTGGTTTTGTGGGCAACATGCTGGTCATCCTCATCCTGATAAACTGCAAAAGGCTGAAGAGCATGA CTGACATCTACCTGCTCAACCTGGCCATCTCTGACCTGTTTTTCCTTCTTACTGTCCCCTTCTGGGCTCACTATGCTGCCGCCCAGTGGGACT TTGGAAATACAATGTGTCAACTCTTGACAGGGCTCTATTTTATAGGCTTCTTCTCTGGAATCTTCTTCATCATCCTCCTGACAATCGATAGGT ACCTGGCTGTCGTCCATGCTGTGTTTGCTTTAAAAGCCAGGACGGTCACCTTTGGGGTGGTGACAAGTGTGATCACTTGGGTGGTGGCTGTG TTTGCGTCTCTCCCAGGAATCATCTTTACCAGATCTCAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCATACAGTCAGTATCA ATTCTGGAAGAATTTCCAGACATTAAAGATAGTCATCTTGGGGCTGGTCCTGCCGCTGCTTGTCATGGTCATCTGCTACTCGGGAATCCTAA AAACTCTGCTTCGGTGTCGAAATGAGAAGAAGAGGCACAGGGCTGTGAGGCTTATCTTCACCATCATGATTGTTTATTTTCTCTTCTGGGCT CCCTACAACATTGTCCTTCTCCTGAACACCTTCCAGGAATTCTTTGGCCTGAATAATTGCAGTAGCTCTAACAGGTTGGACCAAGCTATGCA

GGTGACAGAGACTCTTGGGATGACGCACTGCTGCATCAACCCCATCATCTATGCCTTTGTCGGGGAGAAGTTCAGAAACTACCTCTTAGTC
TTCTTCCAAAAGCACATTGCCAAACGCTTCTGCAAATGCTGTTCTATTTTCCAGCAAGAGGCTCCCGAGCGAGCAAGCTCAGTTTACACCCG
ATCCACTGGGGAGCAGGAAATATCTGTGGGCTTGTGACACGGACTCAAGTGGGCTGGTGACCCAGTCAGAGTTGTGCACATGGCTTAGTTT
TCATACACAGCCTGGGCTGGGGGTGGGGTGGGAGAGGTCTTTTTTAAAAGGAAGTTACTGTTATAGAGGGTCTAAGATTCATCCATTTATT
TGGCATCTGTTTAAAGTAGATTAGATCTTTTAAGCCCATCAATTATAGAAAGCCAAATCAAAATATGTTGATGAAAAATAGCAACCTTTTTA
TCTCCCCTTCACATGCATCAAGTTATTGACAAACTCTCCCTTCACTCCGAAAGTTCCTTATGTATATTTAAAAGAAAGCCTCAGAGAATTGC
TGATTCTTGAGTTTAGTGATCTGAACAGAAATACCAAAATTATTTCAGAAATGTACAACTTTTTACCTAGTACAAGGCAACATATAGGTTGT
AAATGTGTTTAAAACAGGTCTTTGTCTTGCTATGGGGAGAAAAGACATGAATATGATTAGTAAAGAAATGACACTTTTCATGTGTGATTTC
CCCTCCAAGGTATGGTTAATAAGTTTCACTGACTTAGAACCAGGCGAGAGACTTGTGGCCTGGGAGAAGCTTCTTAAATGAGA
AGGAATTTGAGTTGGATCATCTATTGCTGGCAAAGACAGAAGACAGAAGCCTCACTGCAAGCACTGCATGGGCAAGCTTGGCTGTAGAAGGAGACAG
AGCTGGTTGGGAAGACATGGGGAGGAAGGACAAGGCTAGATCATGAAGAACCTTGACGGCATTGCTCCGTCTAAGTCATGAGCTGAGCAG
GGAGATCCTGGTTGTGTTGCAGAAGGTTTACTCTGTGGCCAAAGGAGGGTCAGGAAGGATGAGCATTTAGGGCAAGGAGACCACCAACA
GCCCTCAGGTCAGGGTGAGGATGGCCTCTGCTAAGCTCAAGGCGTGAGGAGGGAGGTGGGAAGGAGGG
AGGTATTCGTGCAGCATATGAGGATGCAGAGTCAGCAGAACTGGGGTGGAAGTTTGGGTTGAAGGTCAGAGGAGTCAGAGAGA
ATCCCTAGTCTTCAAGCAGATTGGAGAAACCCTTGAAAAGACATCAAGCACAGAAGGAGGAGGAGGTTTAGGTCAAGAAGAAGATG
GATTGGTGTAAAAGGATGGGTCTGGTTTGCAGAGCTTGAACACAGTCTCACCCAGGCTGTCTTTCACTGAATGCTTCTGACTTCA
TAGATTTCCTTCCCATCCCAGCTGAAATACTGAGGGGTCTCCAGGAGGAGACTAGATTTATGAATACACGAGGTATGAGGTCTAGGAACAT
ACTTCAGCTCACACATGAGATCTAGGTGAGGATTGATTACCTAGTAGTCATTTCATGGGTTGTTGGGAGGATTCTATGAGGCAACCACAGG
CAGCATTTAGCACATACTACACATTCAAATAAGCATCAAACTCTTAGTTACTCATTCAGGGATAGCACTGAGCAAAGCATTGAGCAAAGGGG
TCCCATAGAGGTGAGGGAAGCCTGAAAAACTAAGATGCTGCCTGCCCAGTGCACACAAGTGTAGGTATCATTTTCTGCATTTAACCGTCAA
TAGGCAAAGGGGGGAAGGGACATATTCATTTGGAAATAAGCTGCCTTGAGCCTTAAAACCCACAAAAGTACAATTTACCAGCCTCCGTATT
TCAGACTGAATGGGGGTGGGGGGGCGCCTTAGGTACTTATTCCAGATGCCTTCTCCAGACAAACCAGAAGCAACAGAAAAAATCGTCTC
TCCCTCCCTTTGAAATGAATATACCCCTTAGTGTTTGGGTATATTCATTTCAAAGGGAGAGAGAGGTTTTTTTCTGTTCTGTCTCATATGA
TTGTGCACATACTTGAGACTGTTTTTGAATTTGGGGGATGGCTAAAACCATCATAGTACAGGTAAGGGAATAGTAAGTGGTGAGAAAC
TACTCAGGGAATGAAGGTGTCAGAATAATAAGAGGTGCTACTGACTTTCTCAGCCTCTGAATATGAACGGTGAGCATTGTGGCTGTCAGCA
GGAAGCAACGAAGGGAAATGTCTTTCCTTTTGCTCTTAAGTTGTGGAGAGTGCAACAGTAGCATAGGACCCTACCCTCTGGGCCAAGTCAA
AGACATTCTGACATCTTAGTATTTGCATATTCTTATGTGAAAGTTACAAATTGCTTGAAAGAAATATGCATCTAATAAAAAACACCT
TCTAAAATAA (**SEQ ID NO: 268**)

Exon 1-2 *SERPINA1*

CTCCTCAGCTTCAGGCACCACCACTGACCTGGGACAGTGAATCgtaagtatgccttccactgcgagaggttctggagaggcttctgagctcccatggcccaggcaggcagcaggt
ctgggcaggaggggggttgtggagtgggtatccgcctgctgaggtgcaggcagatggagaggctgcagctgagcctctatttcataataacagcagccatgagggttgtgtcctgtttcccagtcctgcccggtccccctcg
gtacctcctggtggatacactggttcctgtaagcagaagtggatgaggtgtctaggtctgcagtcctggcacccaggatgggggacaccagccaagccaacagcaacaagcgcagccatttcttctgtttgcacagct
cctctgtctgtcgggggctcctgtctgttgtctcctataagcctcaccacctctcctactgcttgggcatgcatctttctccccttctatagatgaggaggttaaggtccagagaggggtgggaggaacgccggctcacattctccatcc
cctccagatatgaccaggaacagacctgtgccaggcctcagcctcacatcaaaatgggcctcccatgcaccgtggacctctgggccctcctgtccagtggaggacaggaagctgtgaggggcactgtcacccagggctcaag
ctggcattcctgaataatgctctgcaccaggccacggctaagctcagtgcgtgattaagcctcataaccctccaaggcagttactagtgtgattcccatttacagatgaggaagatgggggacagagaggtgaataactggccccaa

atcacacaccatccataattcgggctcaggcacctggctccagtccccaaactcttgaacctggccctagtgtcactgtttctcttgggtctcaggcgctggatggggaacaggaaacctgggctggacttgaggcctctctgatgctc
ggtgacttcagacagttgctcaacctctctgttctcttgggcaaaacatgataacctttgacttctgtcccctcccctcaccccacccgaccttgatctctgaagtgttggaaggatttaatttttcctgcactgagttttggagacaggtcaa
aaagatgaccaaggccaaggtggccagtttcctatagaacgcctctaaaagacctgcagcaatagcagcaagaactggtattctcgagaacttgctgcgcagcaggcacttcttggcattttatgtgtatttaatttcacaatagctctat
gacaaagtccacctttctcatctccaggaaactgaggttcagagaggttaagtaacttgtccaaggtcacacagctaatagcaagttgacgtggagcaatctggcctcagagcctttaattttagccacagactgatgctcccctcttcat
ttagccaggctgcctctgaagttttctgattcaagacttctggcttcagctttgtacacagagatgattcaatgtcaggttttggagtgaaatctgtttaatcccagacaaaacatttaggattacatctcagttttgtaagcaagtagctctgtg
atttttagtgagttatttaatgctctttggggctcaatttttctatctataaaatagggctaataatttgcaccttatagggtaagctttgaggacagattagatgatacggtgcctgtaaaacaccaggtgttagtaagtgtggcaatgatggtg
acgctgaggctgatgtttgcttagcatagggttaggcagctggcaggcagtaaacagttggataatttaatggaaaatttgccaaactcagatgctgttcactgctgagcaggagccccttcctgctgaaatggtcctggggagtgcag
caggctctccgggaagaaatctaccatctctcgggcaggagctcaacctgtgtgcaggtacagggagggcttcctcacctggtgcccactcatgcattacgtcagttattcctcatccctgtccaaaggattcttttctccattgtacagc
tatgaagctagtgctcaaagaagtgaagtcatttaccccaggccccctgccagtaagtgacagggcctggtcacacttgggtttatttattgcccagttcaacaggttgtttgaccataggcgagattctcttccctgcaccctgccgggt
tgctcttggtcccttattttatgctcccgggtagaaatggtgtgagattaggcagggagtggctcgcttccctgtccctggccccgcaaagagtgctcccacctgccccgatcccagaaatgtcaccatgaagccttcattcttttggttta
aagcttggcctcagtgtccgtacaccatggggtacttggccagatggcgactttctcctctccagtcgccctcccaggcactagcttttaggagtgcagggtgctgcctctgatagaagggccaggagagagcaggttttggagtcct
gatgttataaggaacagcttgggaggcataatgaacccaacatgatgcttgagaccaatgtcacagcccaattctgacattcatcatctgagatctgaggacacagctgtctcagttcatgatctgagtgctgggaaagccaagacttg
ttccagctttgtcactgacttgctgtatagcctcaacaaggccctgaccctctctgggcttcaaactcttcactgtgaaaggaggaaaccagagtaggtgatgtgacaccaggaaagatggatgggtgtgggggaatgtgctcctccc
agctgtcaccccctcgccaccctccctgcaccagcctctccacctcctttgagcccagaattcccctgtctaggagggcacctgtctcatgcctagccatgggaattctccatctgttttgctacattgaacccagatgccattctaacca
agaatcctggctgggtgcaggggctctcgcctgtaaccccagcactttgggaggccaaggcaggcggatcaagaggtcaggagttcaagacctgcctggccaacacggtgaaacctcagctctactaaaaatacaaaaattagc
caggcgtggtggcacacgcctgtaatcccagctatttgggaagctgagacagaagaatttcttgaacccgggaggtggaggtttcagtgagccgagatcacgccactgcactccaccctggcagataaagcgagactctgtctcaa
aaaaaacccaaaaacctatgttagtgtacagagggccccagtgaagtcttctcccagccccactttgcacaactggggagagtgaggccccaggaccagaggattcttgctaaaggccaagtggatagtgatggccctgccaggg
ctagaagccacaacctctggccctgaggccactcagcatatttagtgtccccaccctgcagaggcccaactccctcctgaccactgagccctgtaatgatggggaatttccataagccatgaaggactgcacaaagttcagttggg
aagtgaaagagaaattaaagggagatggaaatatacagcactaattttagcaccgtctttagttctaacaacactagctagctgaagaaaaatacaaacatgtattatgtaatgtgtggtctgttccatttggattacttagaggcacgagg
gccaggagaaaggtggtggagagaaaccagctttgcacttcatttgttgctttattggaaggaaacttttaaaagtccaaggggggttgaagaatctcaatatttgttatttccagctttttttctccagtttttcatttcccaaattcaaggacac
ctttttctttgtattttgttaagatgatggttttggttttgtgactagtagttaacaatgtggctgccgggcatattctcctcagctaggacctcagttttcccatctgtgaagacggcaggttctacctagggggctgcaggctggtggtccga
agcctgggcatatctggagtagaaggatcactgtggggcagggcaggttctgtgttgctgtggatgacgttgactttgaccattgctcggcagagcctgctctcgctggttcagccacaggccccaccactccctattgtctcagccc
cgggtatgaaacatgtattcctcactggcctatcacctgaagcctttgaatttgcaacacctgccaacccctccctcaaaagagttgccctctcagatccttttgatgtaaggtttggtgttgagacttatttcactaaattctcatacataaac
atcactttatgtatgaggcaaaatgaggaccagggagatgaatgacttgtcctggctcatacacctggaaagtgacagagtcagattagatcccaggtctatctgaagttaaaagaggtgtcttttcacttcccacctcctccatctacttt
aaagcagcacaaacccctgctttcaaggagagatgagcgtctctaaagcccctgacagcaagagcccagaactgggacaccattagtgacccagacggcaggtaagctgactgcaggagcatcagcctattcttgtgtctgggac
cacagagcattgtggggacagccccgtctcttgggaaaaaaaccctaagggctgaggatccttgtgagtgttgggtgggaacagctcccaggaggtttaatcacagcccctccatgctctctagctgttgccattgtgcaagatgcat
ttcccttctgtgcagcagtttccctggccactaaatagtgggattagatagaagccctccaagggcttccagcttgacatgattcttgattctgatctggcccgattcctggataatcgtgggcaggcccattcctcttcttgtgcctcattttc
ttcttttgtaaaacaatggctgtaccatttgcatcttagggtcattgcagatgtaagtgttgctgtccagagcctgggtgcaggacctagatgtaggattctggttctgctacttcctcagtgacattgaatagctgacctaatctctctggcttt
ggtttcttcatctgtaaaagaaggatattagcattagcacctcacgggattgttacaagaaagcaatgaattaacacatgtgagcacggagaacagtgcttggcatatggtaagcactacgtacattttgctattcttctgattctttcagtgt
tactgatgtcggcaagtacttggcacaggctggtttaataatccctaggcacttccacgtggtgtcaatccctgatcactgggagtcatcatgtgccttgactcggggcctggcccccccatctctgtcttgcag*GACAATGCCG*
TCTTCTGTCTCGTGGGGCATCCTCCTGCTGGCAGGCCTGTGCTGCCTGGTCCCTGTCTCCCTGGCTGAGGATCCCCAGGGAGATGCTGCCCAGAA
GACAGATACATCCCACCATGATCAGGATCACCCAACCTTCAACAAGATCACCCCCAACCTGGCTGAGTTCGCCTTCAGCCTATACCGCCAGCTGGCA
CACCAGTCCAACAGCACCAATATCTTCTTCTCCCCAGTGAGCATCGCTACAGCCTTTGCAATGCTCTCCCTGGGGACCAAGGCTGACACTCACGATG
AAATCCTGGAGGGCCTGAATTTCAACCTCACGGAGATTCCGGAGGCTCAGATCCATGAAGGCTTCCAGGAACTCCTCCGTACCCTCAACCAGCCAG
ACAGCCAGCTCCAGCTGACCACCGGCAATGGCCTGTTCCTCAGCGAGGGCCTGAAGCTAGTGGATAAGTTTTTGGAGGATGTTAAAAAGTTGTACCA

CTCAGAAGCCTTCACTGTCAACTTCGGGACACCGAAGAGGCCAAGAAACAGATTACCGTGGAGAAGGGTACTCAAGGGAAAATTGTGGA TTTGGTCAAGGAGCTTGACAGAGACACAGTTTTTGCTCTGGTGAATTACATCTTCTTTAAAG (**SEQ ID NO: 1795**)
Exon 1-2 *ALB*
CTAGCTTTTCTCTTCTGTCAACCCCACACGCCTTTGGCACAATGAAGTGGGTAACCTTTATTTCCCTTCTTTTTCTCTTTAGCTCGGC TTATTCCAGGGGGTGTGTTTCGTCGAGATGCACgtaagaaatccattttctattgttcaacttttattctatttcccagtaaaataaagttttagtaaactctgcatctttaaagaattatttggcatttatttc taaaatggcatagtatttgttattgtgaagtcttacaaggttatctattaataaattcaaacatcctaggtaaaaaaaaaaaaaggtcagaattgtttagtgactgtaatttctttggcgcactaaggaaagtgcaaagtaacttagagtga ctgaaacttcacagaataggggttgaagattgaattcataactatcccaaagacctatgcttatttaaaaaccacaaaacctgtgctgttgatctcataaaatagaacttgtatttaatatttcatttagtctgtcttctggt tgctgttgatagacactaaaagagtattagatattatctaagtttgaatataaggctataaatatttaaaatagtattcttgtttaaaggcagaggaaataattgaacatcatcctgagttttctgtag gaatcagagcccaatattttgaaacaaatgcataatctaagtcaaatggaaagaaatataaaaagtaacattattactcttgtttctcatgatatttaacaatcctttttttctccctgcccag*ACAAGAGTGAGGTTGCTCA TCGGTTTAAAGATTTGGGAGAAGAAAATTTCAAAGCCTT* (**SEQ ID NO: 1796**)
Exon 1-2 of *Angptl3*
AGAAGAAAACAGTTCCACGTTGCTTGAAATTGAAAATCAAGATAAAAATGTTCACAATTAAGCTCCTTCTTTTTATTGTTCCTCTAG TTATTTCCTCCAGAATTGATCAAGACAATTCATCATTTGATTCTCTATCTCCAGAGCCAAAATCAAGATTTGCTATGTTAGACGATG TAAAAATTTTAGCCAATGGCCTCCTTCAGTTGGGACATGGTCTTAAAGACTTTGTCCATAAGACGAAGGGCCAAATTAATGACATAT TTCAAAAACTCAACATATTTGATCAGTCTTTTTATGATCTATCGCTGCAAACCAGTGAAATCAAAGAAGAAGGAACTGAGAA GAACTACATATAAACTACAAGTCAAAAATGAAGAGGTAAAGAATATGTCACTTGAACTCAACTCAAAACTTGAAAGCCTCCTAGAA GAAAAAATTCTACTTCAACAAAAAGTGAAATATTTAGAAGAGCAACTAACTTAATTCAAAATCAACCTGAAACTCCAGAACAC CCAGAAGTAACTTCACTTAAAgtaagtagaaaataaagaggttcatgtttatgtttcaatggatctttaaaaaaaatatttctaaggcatgccatttgaaatacttgttgcattgttgaaatactttttttccaag aaaaataatctccagaaataaaaatttcctattataatttcaagttagttttttgttccctaatgttatatatgaaaactgaaaatttgcatttatatgaaaattacaaatcggttaaattatacaatctagaacactatgtcattacactattgta aattactgaaggtaagtaaaaaagttaaaaaaaaatttaaaactattctccagtgtttaaaacagattaaatataatacagtaaatggaaaagatttattcatatgaaaatatgctgggcttttctttaattgaagttcagaaaatcaaatttcaaatttagaga tagtacaatttaaataaaatgttaaggacaaaaaatatgtgctatttgaaagaagcatacaaggggaaggaattgccaatattcattttcaaatccattattagtttaaaaatttagattatgatagtgttacaggaaattaatagaaaagaaag aggaaagcaacttataaccaacctactctctatatccag*ACTTTTGTAGAAAACACAAGATAATAGCATCAAAGACCTTCTCCAGACCGTGGAAGACCAATATAAACAATTAAA CCAACAGCATAGTCAAATAAAGAAATAGAAAATCAG* (SEQ ID NO: 1797**)
Exon 1-2 of *AAVS1* (PPP1R12C)
**ATGTCCGGAGAGGATGGCCCGGCGGCTGGCCCGGGGGCGGCTGCCCGGGAGCGGCGACGGGAGCAGCTGCGGCA GTGGGGGGCGCGGCGGGGCGGCCGAGCCTGGCCCCGGAGAGCGCCGCCCGCACCGTCCGCTTCGAGCGGCGCCGGCCGAGTTCCT GGCGGCCTGTGCGGGCGGCGACCTGGACGAGGCGCGTCTGATGCTGCCGCCCGGCCCCGGCGCCGAGCTCGACCC CGCCGCGCCGCCGCCCGCCCGTGCTGGACTCCACCAACGCCGACGGTATCAGCGCCCTGCACCAGgtcagcgccccgcccggcgtctc ccgggccaggtccacctctgctgcgccacctggggcatcctcctcccgttgccagtctcgatccgcccgtgttcctggccctgggctttgccacctatgctgacaccccgtccccttaccattccccttcgacca cccccacttcgaattggagccgcttcaactggccctgggcttagccactctgtgctgaccactgcccccaggcctccttaccattccccttcgaactactctcttccgcattggagtcgcctggcttgcagcctgtg ctgaccatgcagtcctcctaccatcctcctcctgacttcccctcttccgatgtttgagccccctcccagccggtcctggacttttgtctcctcccctgcctgccctgaacctgagccagctcccatagctcagtctggtctatctgcct ggccctggccatttgcacttttgcgctgcctcctctcgcccccgagtgccctgcctcttaacgctgccgtctctctcctgagtccggaccacttttgagcttctactggcttctgcgccgcctctggcccac tgtttccctcctccaggcaggtcctgcttttctgacctgcattctctccctgggcctgtgccgctttctgtctgcagcttgtggcctgggtcacctctacggctggcccagatcctccctgccgcctcctgcagttccgtctctccac tcctctccccttgctctctgctgttgctgcccaaggatgctcttccggagcactccttcctcggcgctgcaccacgtgatgtcctctgagcggatcctcccgtgtctggggtcctctcccgggcatctctcctcctcaccaaccc atgccgtcttcactcgctgggttcccttttccttctcctcctgggggcctgtgccatctctcgtttcttaggatggccttctccggtccccgcctcccccttcttgtaggcctgcatcatcaccgttttctggacaaacccc

aaagtaccccgtctccctggctttagccacctctccatcctcttgctttctttgcctggacaccccgttctcctgtggattcgggtcacctctcactcctttcatttgggcagctcccctaccccccttacctctctagtctgtgctagctcttcc
agccccctgtcatggcatcttccaggggtccgagagctcagctagtcttcttcctccaacccgggcccctatgtccacttcaggacagcatgtttgctgcctccagggatcctgtgtccccgagctgggaccaccttatattcccaggg
ccggttaatgtggctctggttctgggtacttttatctgtcccctccaccccacagtggggccactagggacaggattggtgacagaaaagccccatccttaggcctcctccttcctagtctcctgatattgggtctaacccccacctcctgt
taggcagattccttatctggtgacacacccccatttcctggagccatctctctccttgccagaacctctaaggtttgcttacgatggagccagagaggatcctgggagggagagcttggcagggggtgggagggaagggggggatg
cgtgacctgcccggttctcagtggccaccctgcgctaccctctcccagaacctgagctgctctgacgcggccgtctggtgcgtttcactgatcctggtgctgcagcttccttacacttcccaagaggagaagcagtttggaaaaacaa
aatcagaataagttggtcctgagttctaactttggctcttcacctttctagtccccaatttatattgttcctccgtgcgtcagttttacctgtgagataaggccagtagccagccccgtcctggcagggctgtggtgaggaggggggtgtcc
gtgtggaaaactccctttgtgagaatggtgcgtcctaggtgttcaccaggtcgtggccgcctctactccctttctctttctccatccttctttccttaaagagtccccagtgctatctgggacatattcctccgcccagagcagggtcccgct
tccctaaggccctgctctgggcttctgggtttgagtccttggcaagcccaggagaggcgctcaggcttccctgtcccccttcctcgtccaccatctcatgcccctggctctcctgccccttccctacaggggttcctggctctgctcttca
gactgagccccgttcccctgcatccccgttcccctgcatccccccttcccctgcatcccccagaggccccaggccacctacttggcctggaccccacgagaggccaccccagccctgtctaccaggctgccttttgggtggattctcc
tccaactgtggggtgactgcttggcaaactcactcttcggggtatcccaggaggcctggagcattggggtgggctggggttcagagaggagggattcccttctcaggttacgtggccaagaagcaggggagctgggtttgggtca
ggtctgggtgtggggtgaccagcttatgctgtttgcccaggacagcctagttttagcactgaaaccctcagtcctaggaaaacagggatggttggtcactgtctctgggtgactcttgattcccggccagtttctccacctggggctgtg
tttctcgtcctgcatccttctccaggcaggtccccaagcatcgccccctgctgtggctgttcccaagttcttagggtaccccacgtgggtttatcaaccacttggtgaggctggtaccctgcccccattcctgcaccccaattgccttagt
ggctagggggttgggggctagagtaggaggggctggagccaggattcttagggctgaacagagaagagctgggggcctgggctcctgggtttgagagaggaggggctggggcctggactcctgggtccgagggaggaggg
gctggggcctggactcctgggtctgagggtggagggactgggggcctggactcctgggtccgagggaggaggggctggggcctggactcgtgggtctgagggaggaggggctgggggcctggacttctgggtcttagggag
gcggggctgggcctggaccccctgggtctgaatggggagaggctgggggcctggactccttcatctgagggcggaagggctggggcctggcctcctgggttgaatggggaggggttgggcctggactctggagtccctggtgc
ccaggcctcaggcatctttcacagggatgcctgtactgggcaggtccttgaaagggaaaggcccattgctctccttgccccccctcccctatcgccatgacaactgggtggaaataaacgagccgagttcatcccgttcccagggcac
gtgcggccccttcacagcccgagtttccatgacctcatgctcttggccctcgtagctccctcccgcctcctccagatgggcagctttggagaggtgagggacttgggggggtaatttatcccgtggatctaggagtttagcttcactcctt
cctcagctccagttcaggtcccggagcccacccagtgtccacaaggcctggggcaagtccctcctccgacccccctggacttcggcttttgtcccccccaagttttggacccctaagggaagaatgagaaacggtggcccgtgtcagc
ccctggctgcagggccccgtgcagagggggcctcagtgaactggagtgtgacagcctggggcccaggcacacaggtgtgcagctgtctcacccctctgggagtcccgcccaggcccctgagtctgtcccagcacagggtggc
cttcctccaccctgcatagccctgggcccacggcttcgttcctgcagagtatctgctggggtggtttccgagcttgacccttggaaggacctggctgggtttaaggcaggaggggctgggggccaggactcctggctctgaaggag
gaggggctggaacctcttccctagtctgagcactggaagcgccacctgtgggtggtgacgggggttttgccgtgtctaacaggtaccatgtggggttcccgcacccagatgagaagccccctcccttccccgttcacttcctgtttgc
agatagccaggagtcctttcgtggtttccactgagcactgaaggcctggccggcctgaccactgggcaaccaggcgtatcttaaacagccagtggccagaggctgttgggtcattttccccactgtcctagcaccgtgtccctggatc
tgttttcgtggctccctctggagtcccgacttgctgggacaccgtggctggggtaggtgcggctgacggctgtttcccacccccag*GCCTGCATTGATGAGAACCTGGAGGTGGTGCGCTTCTTGG*
TGGAGCAGGGCGCCACTGTGAACCAGGCAGACAACGAGGGCTGGACGCCACTGCACGTGGCCGCCTCCTGTGGCTACCTAGATATCGCCAG
(SEQ ID NO: 1798)

***AAVS1* (alternative; GenBank Accession No: S51329.1)**

GAATTCCTAACTGCCCCGGGGCAGTCTGCTATTCATCCCCTTTACGCGGTGCTACACACACTTGCTAGTATGCCGTGGGGACCCCTCCGGCC
TGTAGACTCCATTTCCCAGCATTCCCCGGAGGAGGCCCTCATCTGGCGATTTCCACTGGGGGCCTCGGAGCTGCGGACTTCCCAGTGTGCAT
CGGGGCACAGCGACTCCTGGAAGTGGCCACTTCTGCTAATGGACTCCATTTCCCAGGCTCCCGCTACCTGCCCAGCACACCCTGGGGCATC
CGTGACGTCAGCAAGCCGGGCGGGGACCGGAGATCCTTGGGGCGGTGGGGGGCCAGCGGCAGTTCCCAGGCGGCCCCCGGGGCGGGCGG
GCGGGCGGGTGGTGGCGGCGGTTGGGGCTCCGGGCGCGTCGCTCGCTCGCTCGCTGGGCGGGCGGGCGGTGCGATGTCCGGAGAGGATGG
CCGGCGGCTGGCCCGGGGGCGGCGGCGCGGCTGCCCGGGAGCGGCGACGGGAGCAGCTGCGGCAGTGGGGCGCGGGCGGGCGCCGAGCC
TGGCCCCGGAGAGCGCCGCGCCCGCACCGTCCGCTTCGAGCGCGCCGCCGAGTTCCTGGCGGCCTGTGCGGGCGGCGACCTGGACGAGGC
GCGTCTGATGCTGCGCGCCGCCGACCCTGGCCCCGGCGCCGGAGCTCGACCCCGCCGGCCGCCGCCCGCCCGCGCCGTGCTGGACTCCACC
AACGCCGACGGTATCAGCGCCCTGCACCAGGTCAGCGCCCCCCGCGGCGTCTCCCGGGGCCAGGTCCACCCTCTGCGCCACCTGGGGCATC

CTCCTTCCCCGTTGCCAGTCTCGATCCGCCCCGTCGTTACTGGCCCTGGGTTTNCACCCTATGCTGACACCCCGTTCCAGTCCCCTTACCATT
CCCTTCGACCACCCCACTTCCGAATTGGAGCGCTTCAACTGGCTGGGCTAGCACTCTGTGTGACACTCTGAAGCTCTACATTCCCTTCGACC
TACTCTCTTCGATTGGAGTCGCTTTAACTGGCCCTGGCTTTGGCAGCCTGTGCTGACCCATCGAGTCCTCCTTACCATCCCTCCTCGACTTC
CCCTCTTCCGATGTTGAGCCCCTCCAGCCGGTCCTGGACTTTGTCTCCTTCCCTGCCCCTCTCCTGAACCTGAGCCAGCTCCCATAGCT
CAGGTCTGGTCTATCTGCCTGGCCCTGGCCCATTGTCACTTTGCGCTGCCCCCGAGTGCCCTTGCTGTGCCGCCGGAACTCTG
CCCTCTAACGCTGCCGTGCCGTCTCTCCTGAGTCCGGACCACTTTGAGCTCTACTGGCTTCTGCGCGCTCTGGCCCACTGTTTCCCCTTC
CCAGGCAGGTCCTGCTTTCTCTGACCAGCATTCTCTCCCCTGGGCCTGCTTTCTGTCTGCAGCTTGTGGCCTGGGTCACCTCTACGGC
TGGCCCAAGATCCTTCCCTGCCGCCTCCTTCAGGTTCCGTCTTCCCCTCTTGCTCTCTGCTGTGTTGCTGCCCAAGGATG
CTCTTTCCGGAGCACTTCCTTCTTCTCGGCGCTGCACCACGTGATGTCCTCTGAGCGGATCCTCCTCTCCGGGCATCTCTCC
TCCCTCACCCAACCCCATGCCGTGTTCACTCGCTGGGTTCCCTTTCCTTCTTCCTTCTGGGGCCTGTGCCATCTCTCGTTTCTTAGGATGGCCT
TCTCCGACGGATGTCTCCCTTGCGTCCCGCCTCCCCTTCTTGTAGGCCTGCATCATCACCGTTTTTCTGGACAACCCCAAAGTACCCCGTCTC
CCTGGCTTAGCACCTCTCCATCCTCTTGCTTTCTTTGCCTGGACACCCCGTTCTCCTGTGGATTCGGGTCACCTCTCACTCCTTTCATTTGGGC
AGCTCCCCTACCCCCCCTTACCTCTCTAGTCTGTGCTAGCTCTTCCAGCCCCCTGTCATGGCATCTTCCAGGGGTCCGAGAGCTCAGCTAGTCT
TCTTCCTCCAACCCGGGCCCTATGTCCACTTCAGGACAGCATGTTTGCTGCCTCCAGGGATCCTGTGTCCCCGAGCTGGGACCACCTTATAT
TCCCAGGGCCGGTTAATGTGGCTCTGGTTCTGGGTACTTTTATCTGTCCCTCCACCCCACAGTGGGGCCACTAGGGACAGGATTGGTGACA
GAAAAGCCCCCATCTCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTCGTCTAACCCCCACCTCCTGTTAGGCAGATTCCTTATCTGGTGACA
CACCCCCATTTCCTGGAGCCATCTCTCTCCTTGCCAGAACCTCTCTAAGGTTTGCTTACGATGGAGCCAGAGAGGATCCTGGGAGGGAGACTT
GGCAGGGGGTGGGAGGGAAGGGGGGATGCGTGACCTGCCGGTTCTCAGTGGCCACCCTGCGCTACCCTCTCCCAGAACCTGAGCTGCT
CTGACGCGGCTGTCTGGTGCGTTTCACTGATCCTGGTGCTGCAGCTTCCTTACACTTCCCAAGAGGAGAAGCAGTTTGGAAAAACAAAATC
AGAATAAGTTGGTCCTGAGTTCTAACTTTGGCTCTTCACCTTTCTAGNCCCCAATTTATATTGTTCCTCCGTGCGTCAGTTTTACCTGTGAGA
TAAGGCCAGTAGCCACCCCCGTCCTGGCAGGGCTGTGGTGAGGAGGGGGGTGTCCGTGTGGAAAACTCCCTTTGTGAGAATGGTGCGTCCT
AGGTGTTCACCAGGTCGTGGCCGCCTCTACTCCCTTCTTCTCCTTCTTTCCTTAAAGAGCCCCCCAGTGCTATCTGGACATA
TTCCTCCGCCCAGAGCAGGGTCCGCTTCCCTAAGGCCCTGCTCTGGGTTTGAGTCCTTGCAAGCCCAGGAGAGCGCTAGCTTCC
CTGTCCCCCTTCCTCGTCCACCATCTCATGCCCTGGCTCTCCTGCCCCTTCCTACAGGGGTTCCTGGCTCTGCTCTTCAGACTGAGCCCCCGTT
CCCCTGCATCCCCGTTCCCCCTGCATCCCCCCTTCCCCCAGAGCCCCAGGCCACCTACTTGGCCTGGAACCCCCACGAGAGGCCAC
CCCAGCCCTGTCTACCAGGCTGACCTTTTGGGTGATTCTCCTCCAACTGTGGGGTGACTGCTTGGGCAAACTCACTCTTCGGGGTATCCCAG
GAGGCCTGGAGCATTGGGGTGGGCTGGGGTTCAGAGAGGAGGGATTCCCTCCAGGTTACGTGGCCAAGAAGCAGGGGAGCTGGGTTTGGG
TCAGGCTGGGTGTGGGGTGACCAGCTTATGCTGTTTGCCCAGGACAGCCTAGTTTTTAGCGCTGAAAACCCTCAGTCCTAGGAAAACAGGGAT
GGTTGGTCACTGTCTCTGGGTGACTCTTGATTCCGGCCAGTTTCTCCACCTGGGGCTGTGTTTCTCGTCCTGCATCCTTCTCCAGGCAGGTC
CCCAAGCATCGCCCCCCCTGGCTGTTCCCAAGTTCTTAGGTACCCCACGTGGGTTTATGAACCACTTGGTGAGGCTGGTACCCTGCCCCCCATT
CCTGCACCCCCAATTGCCTTAGTGGCTAGGGGGTTGGGGGCTAGAGTAGGAGGGGCTGGAGCCAGGATTCTTAGGGCTGAACACAGAGCCGAG
CTGGGGGCCTGGGCTCCTGGGTTTGAGAGAGGAGGGCTGGGGCCTGGACTCCTGGGTCCGAGGGAGGGGCTGGGGCCTGGACTCCT
GGGTCTGAGGGTGGAGGGACTGGGGGCCTGGACTCCTGGGTCCGAGGGAGGGGCTGGGGCCTGGGTCTGAGGGAGGAG
GGGTCGGGGGCCTGGACTTCTGGGTCTTAGGGAGGCGGGGCTGGGCCTGGACCCCCTGGGTCTGAATGGGGAGAGGCTGGGGCCTGGACT

CCTTCATCTGAGGGCGGAAGGGCTGGGGCCTGGCCTCCTGGGTTGAATGGGGAGGGGTTGGGCCTGGACTCTGGAGTCCCTGGTGCCCAGG
CCTCAGGCATCTTTCACAGGGATGCCTGTAC (**SEQ ID NO: 267**)

Exon 1-2 of *ApoC3*

CTGCTCAGTTCATCCCTAGAGGCAGCTGCTCCAGgtaatgccctctgggaggggaaagaggaggggaggaggatgaagaggggcaagaggagctccctgcccagcccagccagcaa
gcctggagaagcacttgctagagctaaggaagcctcggagctgacgggtgccccccacccctcatcataacctgaagaacatggaggcccgggagggtgtcacttgcccaaagctacacagggggtgggctggaagtgg
ctccaagtgcaggttcccccctcattcttcaggcttagggctggaggaagccttagacagcccagtcctacccagacaggaaactgaggcctggagaggccagaaatcacccaaagacacacagcatgttggctggactgg
acggagatcagtccagaccgcaggtgccttgatgttcagtctggtgggttttctgctccatcccacccacctcccttgggcctcgatccctcgccctcaccagtcccccttctgagagcccgtattagcagggagccggcccctact
ccttctggcagacccagctaaggttctacttagggccacgccacctcccccaggagggtccagaggcatgggacctgggtgcccctcacaggacacttccttgcag*GAACAGAGGTGCCATGCAGCC*
CCGGGTACTCCTTGTTGTTGCCCTCCTGGCGCTCCTGGCCTCTGCCC (**SEQ ID NO: 1799**)

Exon 1-2 of *ASGR2*

TCCCAGGCCAGGAGGTGCCAGGTTAGAGCTTGGACACTCAGACCCTCCAGTGCAGGCCTTGCCCGGACCAGCCCCTCCTCCACTC
TCCTCTCCCCTGTGAGCTCCACCTGCCCCAGTTCTCCTGGCTTTAACCCCCTCCTTGGCCAAGGCCAGGGTTGCCTGCGGGAGCCAG
GTGTCCGCTCTCCACACCTTTCACAGgtaactgaggctgggggaggcaactgggacgaggatgcttgggaacaaggaggccttggttcagggatgggggaggacccatgggaagctgtgggtggtc
gggtccatgagcgtgggggaggggtctcctggccaggtcaaggcccaagcggtctcctgagaaggaaaggacccagcgggatttcaaacctcatcccaaacgccctgccaacctcagatcctacccaggccaccaccccgaa
ccctccgaccctgcgttccatcccaccaccatcaacctccacgtgaatctgcacctctgccccag*CCCCAGCCCTCAGAGCAACCTCAGCCCAGCCCAGCTCCAGCTCCAGCTC*
CAGCCCGGGCCCCATCATGGCCAAGGACTTTCAAGATATCCAGCAGCTGAGCTCGGAGGAGAAAATGACCATCCTTTCCATCAAGGTGAGGGCCAGG
CACTCGCAGGCTGAATCCCAGGAGAGGAAATCCATTTTTGAAAG (**SEQ ID NO: 1800**)

Exon 1-2 of *CCR5*

CTTCAGATAGATTATATCTGGAGTGAAGAATCCTGCCACCTATGTATCTGGCATAGTgtgagtcctcataaatgcttactggtttgaagggcaacaaaatagtgaaca
gagtgaaaatcccccactaagatcctgggtccagaaaaagatgggaaacctgttagctcacccgtgagcccatagttaaaactctttagacacaacaggttgtttccgtttacagagaacaatatattggtggtgagcatctgtgtgggg
gttgggggtggataggggatacgggggagagtggagaaaaagggggacacagggttaatgtgaagtccaggatccccccctcacatttaaagttggctttaagttggcttttaattaatagcaacttcttaagatttcttaaccttt
agccttactgttgaaaagccctgtgatcttgtacaaatcatttgctcttcttggatagtaatttctttactaaaatgtgggcttttgactagatgaatgtaaatgttcttctagctctgatatcctttatattttctaacag*ATTCTGTGT*
AGTGGGATGAGCAGAGAACAAAAACAAAATAATCCAGTGAGAAAAGCCCGTAAATAAACCTTCAGACCAGAGATCTATTCTCTAGCTTATTTTAAGCTC
AACTTAAAAGAAGAACTGTTCTCTGATTCTTTTCGCCTTCAATACACTTAATGATTTAACTCCACCCTCCTTCAAAGAAACAGCATTTCCTACTTTTAT
ACTGTCTATATGATTGATTTGCACAGCTCATCTGGCCAGAAGAGCTGAGACATCCGTTCCCCCTACAAGAAACTCTCCCCCG (**SEQ ID NO: 1801**)

Exon 1-2 of *FIX* (F9)

ACTTTCACAATCTGCTAGCAAAGGTTATGCAGCGCGTGAACATGATGGCAGAATCACCAGGCCTCATCACCATCTGCCTTTTA
GGATATCTACTCAGTGCTGAATGTACAGgtttgttccttttttaaaatacattgagtatgcttgccttttagatatagaaatatctgatgctgtcttcttcactaaatttgattacatgatttgacagcaatattga
agagtctaacagccagcacgcaggttggtaagtactggttcttgttagctaggtttcttcttcatttttaaaactaaatagatcgacaatgcttatgatgcatttatgtttaataaaacactgttcagttcatgtaattcctgtt
agaaaacattcatctccttggtttaaaaaaattaaaagtgggaaaacaaagaaatagcagaatatagtgaaaaaaaataaccacattatttttgtttggactaccacttgaaatcaaaatgggaaacaaaacaatggccttat
ttacacaaaaagtctgatttaagatatatgacatttcaaggtttcagaagtatgtaatgaggtgtgtctcttaatttttaaatttcatttaaagttaccttcattagcaagctgttagttatcaccaaa
gcttttcatggattaggaaaaaatcatttgtctctatgtcaaacatcttggagttgatatttgggaaacacaatactcagttgagttccctagggagaaaagcaagcttaagaattgacatagaagtaggaagttagctaatgcaaca
tatatcactttgttttttcacaactacagtgacttattgtatttcccagaggaaggcatacagggaagaaattatcccatttggacaaaacagcatgttctcacaggaagcatttatcacacttactgtcaacttctagaatcaaatctagtagc
tgacagtaccaggatcaggggtgccaacccctaagcaccccagaaagctgactggccctgttggttcccactccagacatgatgtcagctgtgaaatcagactgaaataacgataaaaaaaaaatacagaggttaaacta
gcaaagtgagtaaagtcaagggataaagaaaatttgttggaaaactcacaaaagcagggacataaagcaaggccatttagatatatctcattagtgtgacatctgggaggacaaagcatccaaaccctttcttctatataagtggtgagatg

atgaaggttgtaagagggcttctgccccccttgaagacttcagatgctggggaaaggatagataagaataaggatgaaccaggcttttggagcctgggaaataatgactagcgataaacctgaagggaagttaagtatacgatcccca
gataatactaaggagaaaggcaatgtgattctgcagccattgtagccagagataataagcccttgaggaaggggccaggggaatttttctaaggatagacagtattaatgcagcactctcttctgctattaaactctcattggcttctaaa
aggagtttcggtgagtgatttgctgagatgttgcattttcatggctgctgcctttaggtattattgcaacagtttggaattttgaaattaaaacagttctgtaaaaccagtttagttttgtaaagtgtatgcatcaaagatgtccttcattcagacat
tactgagttacaactatggtgccaggtactgtgtcagggtactaggggtatggggataaaccagactccctctttgatctaaagcagcatgaggccaggtgagaggtttcaatataatgtgataaaatgtgcactaggtactaagggat
catagagaaaggaacacattaaatggggaaacaattgatagagagagaatattttcatctgggtcttaaaagatgagtaggcgttctctctctttaaatgtctgatataagggcattttatgcaaagaggatcactcgtgcaaagactcag
atttgcaagaacgtgaggtatttcaggagttttgtatggttccatatggactatgacaagtgagacaggtaaactaggcagagctggtcatcagataatgaagtcattaacctaaggagattggacaataaaatgcaatatggaggtatc
gaagtataaacataaggagtaccactgatggctgatttaggatgcccagtctggcaacacgctaatgaaatgatagtgggggagggggccgtaccaagactaggagagagcagtcctgagactattgcaattatctgcgggagac
ataaaggctagaacctgaacagtagcagtacaaaaaaagaggggagttcaaatgatattaaggaagtagaagtggtatgacttaaccatctgggtatggaaggggaaatggctagagtcttggggactttgtgtttgatgtgattatgg
accacagaataatgtctaagagaactggctctttagtctgactgccagagtctgaatcctgaatgttttagtatgttaccttgggcaaagcccttagcctctatgaatctatcttcctcattcataaaaataagatgacagtgcctatctcgtg
ggacttttgtgaggatgaagtgagataatggatgcaaagttactgagcacagtgtccaacacagcagaagcattacatatacattagctattactggctacattatgatatacagttagggagttggaaagataatctgaaattcaggag
acgtatctgactataggtgagtatttggaactcattgttctgtaaacagtagttacagccacgtgtgtgggcatctggagagtgagcatggatattgtgatacctagtacagtgcctggcagtagtggttgtatgctcagtaaattttgttga
cagggtcagggccggactagactgtggtaagcaaggcctgtagggcataaatatacttgtatgccccgagaagtgaggacctcttaaatattgtgccctacatgccttgtttggttcactcttgtcccagccctagcaagtatatataag
gtgaaaaaggaaaaagctgaggctggagcctgggagaaccctggacatttaagggccatggagaggaacaggagttaatcaattcaagtgctggatggataacaggagttagagcaaagcggggaaccagaatagagtgatta
ttataaaaagagtttcctaaaaagggagagatcaacaattagaaattatttagagcagccagtaaatacataaactcaaaattattctttaggtcattcctgattgtgacaatagtcatttcattatataaatgtgattaaggaaggaaaagag
ctacacagaagttattaaagagctaaagagaattgagaaatttaaaacagaagaaagtagggccaacatgaaaggagtagggagaaaaagagataaccagcatattgtctcactgatcctgccaacacctgtgagatagatattctta
ttatactactgttaaacctaatttacatttgacaaagttaaggttcagagcttgtgtgacttgtccaaggtcacaggtctagaggaggcagatacttgattcaaacctatttctgtctgatctgattctaaagtctgtttttttcactcaaccacact
gtacagtcagctctccttgtgagttccacagccacagattcaattaactgcagatcaaaaatattcaagaaaaaaatggatggttgcatctctactgaacatgtacagactcttttatctttcattattccctaaacaatacagcataacaact
atttacatagcatttacattgtattagctattaagagaaacctagagatgatttaaagtacaaaggaggatgtgtttaggttatatgcaaatagtaagccatttttatatcggagacttgagcatccacagatcttgatatttgcagggggtctt
gccaccaattttccatggatactgaggaacgactgtaaatggatgcaggcatggatgctatttaggagtgtccagggccaagtaaatgagttgctgagcagagaggtgggtggaggctgtgaggcatcaatatgtggtggcatccat
ctgcattttggtgatttttttccttcacagtcctcggctgtctgggaagagaaggatgaaggcagatggctgctccaatttaggggctaggattgcagggtgggcacagcattgcaaacgagtgaaggaaattgagaaatatggccaat
gaagagttgaagagaggcctggcatggtggctcacacctataatcccagcactttcagaggcccaggcaggcagatcacttgaggtcaggagttcgacaccagcctggccaacaaggtgaaatggtgaaaccccggctttactaa
aaatacaaaaattagctgggcatggtggcgggtgcctgtaatcccagctacttgggaggctgaggcaggagaatagcttgaacctgggagatggaggttgcagtgagctgagatcgcaccactgcactccagcctgggcgacag
agcaagactctgtcaaaaaaaaaaagagttgaagagaaaaagtctaggctaaattcaaagaaaaaaagtgagcccaaaaggaacttgcagagcaagggaaaagcagggatgtcaagggactagaacactccataaagtgaacag
ctgcaatgaaaataagggaagaaagtttagttcatctccgtttctttcctttcctttttactttcctttctcttcctttttggagttaatcaggaagtagtcccaaataccccagaaagttcatcttataagcccttggtcctcttgagatggtatcag
atatattgctagacccttgaagaaaggaacaactccaggcaacttcttgagtccctgttattaattttatacatacacacacatatatgtatatacatgaaaacacacaaacacatgtgtgtgtatacagccatgcattccttaacaatgggg
atatattctgagaaatgtgtcattaagcaatttcatcattgtgcgaacataatagagtgtacttacctaaacctaaatggtatagcttactacatacctaggttgtattgatgtggcctattgctcctaggctcctgggctgcaaacctgtacag
catgtgactgtactgaacactgtaggcaatggtaacagtggtatttgtgtatctaaacatagaaaaggtacagtgaaaatacagtattataaccttatgggaccactgtcgtataatgtggtccatcattgaccaaaatgtcattgtgcagc
aaatgattatctcatatatatatatatatgatatgatatatatgatatatatgtgtgtgtatatatgtatacatatatatgtgtacatatatgtatacatatatacacacacatatatatgtacacacatatatgtatatatatgtacacacatatatgtat
atatatatgtacacacacacacatagagagagagagagaggagaggagaggaaggagggagggaaggagaaatatgattcagatagagacatctatcctccagagttcaggagtgtctcttcagactaggtagatgtagcttaaa
aaaaacatatcctgaattctagagagatgcttaaatcactgcaattcctataacacttgccaaccaaaggtgctgttgatctgaaattgctttttaaattaatgcagtgatttttctttaacatctagtgacagacactggggtcacatttgcag
ctggaccataattaggcttctgttcttcaggagacatttgttcaaagtcatttgggcaaccatattctgaaaacagcccagccagggtgatggatcactttgcaaagatcctcaatgagctattttcaagtgatgacaaagtgtgaagttaa
ccgctcatttgagaactttctttttcatccaaagtaaattcaaatatgattagaaatctgaccttttattactggaattctcttgactaaaagtaaaattgaattttaattcctaaatctccatgtgtatacagtactgtgggaacatcacagattttg
gctccatgccctaaagagaaattggctttcagattatttggattaaaaacaaagactttcttaagagatgtaaaattttcatgatgttttctttttgctaaaactaaagaattattcttttacatttcag*TTTTTCTTGATCATGAAAA
CGCCAACAAAATTCTGAATCGGCCAAAGAGGTATAATTCAGGTAAATTGGAAGAGTTTGTTCAAGGGAACCTTGAGAGAGAATGTATGGAAGAAAAGT
GTAGTTTTGAAGAAGCACGAGAAGTTTTTGAAAACACTGAAAGAACA* **(SEQ ID NO: 1802)**

Exon 1-2 of *Gys2*

AGATACTGACAGGGCAGATACCGTCCTCACAATACCTGCCCAGAAAGACGAGAAAGAGGAGGAAGAATTCCTCCTTCCACCAGGA
ATTCTGTGGGAAGCACATAAGATTTCATGCTACTAGTTTATTCCCAAGAGAAGCTACCAAAGCCTGGTAACTCTACCAACTCTAACT
TTTGTGCCTGTAAGTTCTCTTCTCCTGGGATTACAACTAATTGAAACAGGAATTCAAAGGAGTCTCGGAGGACTGTAAGAAGAATG
CTTCGAGGCCGATCCCTCTCTGTAACATCCCTGGGTGGGCTTCCCCAGTGGGAAGTCGAAGAACTTCCTGTGGAGGAGTTACTGCT
CTTTGAAGTTGCTTGGGAAGTGACCAATAAAGgtttgtactgctcctgaaggtggatcagtacaccttcactaggtgactgaggaatgtagagtaagatttcacagataactctctagtgataacca
attgctagtgagtgaggaaagcatttaaacggaatatatatttaggctacatttaaccttgtctattacaacaggctatgaatacagacttgtgtatgtgttgaagagggagaaagttgaaaatattagccattgccatggaatgtgcatttaa
attaaaatttaaaatccatattatgacaacatagaaataaaaagttaatatttacatttttttgttttaaaatttttctcagcaaaatagtataattaaacattttattttaggaagatgtatatgtgattagaaccagttggaaacctcaactctacattc
tgttagtaattatatgaaatgcttaagaatgtttcatattcacaattagcttggaattaatgttttaattcaggagcttctcaaaaggcaaaatactgtgctcctttcagaatgaaagtaattttttttcttcaatgctatgcttatgtaggaattataaa
ataaagaaaatcaaattgataaaacaatttgcgagtgaggtataacctcatctgtgctgggtaagagaaggcaaggtagagaaccacaatgtaaatgattttgaaaataaagacatctgtaaaattataaaacaattccttgtgggtttata
ctctctagtttaaaagtgtcaagactattgaaggaacattagcaaagaaaaagtcatggttttgaaggctgagttcctcaagatcttgaagaatcaagccaactgattgctgggtaaatggcagtgatgacttgtgtgatttaaacgatctg
aaattatgtcatgtgttgtgatatataacacatttattttcattttgtgtttactatttcatgtctattacagtaatctaaaggatctgaaataataaccagaatgtgatatattacttccttcttttggctaatcagagccagcataatctcaaaattaa
atgcattatatatggcatgcattctgaaaaatcagaattcaccttatactaactgttcaaagttctcgtgtgtatgcattcataaaatcaaaattaaaagtcaattttgcctatgtagaagtggacagaatgcatataagagaacaaagcacgg
tcttattctctgtagtttagtgagaaagaatctaggaggacagaaaccttattacagagaagtgaagtattttatgtgtacctgtttctgttgtttctttttaaattattttgtattcttttcctcagagatctttatcctgaaatgagcatgtgtttttaat
gtcctttttctagtcacaaatttttccagtccagtttttaaatatgcccaaagttcctgggaacagaaccatctctatcacatgtgaatcccggagtatgccttgaaataaaagttttctctagagagtgggcaggagctcaaggacatagctt
attggtggataaaactggaagggacttgttagcactgggaatcagtccagaagatctctgcacaagttcctgcacctgagcagagccccttctgctttcctcttttctcatcatttaacaaacattaatggagcactgttacctgggcactg
ggaacacgccagtgaaactaagttattttatggatcttataatcagtagcatgtgtgttgatctcagggtctagaacagtgcctggtacatagaaaatgctcaatggatttttttttgaaaaaatttgtttttcaacccctaagctgcctttaatac
ttattttctatcatttccctttcaaggggaagggaagtgtttttgtggagaggtccccacaatgtgatttacacaagttgatatcatcttcaatgtatctttattttaatattatcatagcctggtatggtttgagattttcatatagatggataattatt
gtgtttattcctaataatacattgccactatagtttctttaaatttaatgatttgaatttttacttccctgcatttcatagtagtgatatgaggtatcttagagatcatgaatcccactcctccagttcatagagcaaattgattcgtagaggtatttaat
gatcaagcaaatggcagagttaggaccaggatttaggtcttttccaaggccaattcactttttttactattatgtaagaattaaatattaaaacctcagttttcaaaggaaatgtcctatagatgaaagaacacagtacttgcagtcaaaggtct
tggatttctgaaactttctgaaaaagaaccccttaaactgaataacagtttcttcatgaataaaatgaatgttatactctccagttatcttaccatgttttctggcaaacaaggaaatgtattcacatccaagaaaacaaaaggtttcttaggag
aagaatcccttgctgccagaattttgcccatgaacaatggaaacgtgaaaaacataacagatgagagagtcacagttagttttgggtactaattctcagacttcaaaataaccttttttaatttattaatattaaccttgaaatttattgcttttatt
cattaatccattttattattcagtaaggattcatcaagtgtctattaggtatcaggcactgagaggcgagggagatagagttaagaatacagaactgattcttaatcttcaagaattgacaaatttgtagaaggtagacattgaacatacaat
acagtcaaatgtgatgagtgtgagactaagggaaccaggtgtcctaagactatacagtgtaccgtggaatggggagtggtttctaacctagttacctagaggtttgggaaaaacctctctaatgatgagatgtttgagctgacacttgaa
ggataaatagctaaaggggagaaagcgttctatgtgagagaactatatatgccaagatgcagaagtgttctctgttcgaagatctgagcggggctcagagagggaaaaggagaggagtatgagagaagatgggagagacctgca
gggactagattgtgaagctcaggggggagttaatagccttaaggagttttgtctccagggtatgaacaatgaatgaggagtcactgaaattctatccacagtagtgggaagtactaagtagtcaggtatgttttggaagaagcactgttac
cagttgagggagaggagaggagggaaattagactgtggtgtgacccaggatggaaagttttgcccaaatgtgggggaaaaatgtttactctctcctttgcccaggtcttactacaatcatttcttctgcacacacttgctcagaacattac
aaactctggagaaattaacatgttaaggatatttaggatcattaaaaaaagtttaataaatgttgtcacacataatgaaagttaatttaagaattgtaaagacttgagcgatgtagtccctaaataaatgcttccctcttaaaatggactagaaac
aatttgtttttgcctaagccaacgtattgctggacataatttacctgcttcctcctcagtcttgacactacacctctgaattacatggtacaagtggcacagggaagaattcagaaaactagagtcaaagtggttatggaagaaattataata
atgagtccactcttctaggacaggagttctaaaggtgaggactttgatataagtgaactcaaggagtttctaaaccttcagaaattatatctaaaaaattaggtgcacgtatatagaagtacttttctctgagaagaggacttcacacttccatc
agcttcctgaaggcccctgtaattgccgtaaagaacaactgctctaaagtttaatactaggctactcggctgtttttttgttaacacagaagttttccaaatggcttttctctagagcaaaatgaaaaggatcccatccttcccgagacctaac
gatgtttatctttcgtgatactgtttttgttcttcctctttttctcctaccccccaccacttttaaagatccatcaaggcccagcatggtgattcacacctgtaatctcaacacttttcaaggctgaggtgggagaattgtttgagaccaagagttg
gataccagcctggggaacatagggagaccctgtttctacagaaatttaaaaaaatttctctagctggccatggaggctcacacctgtagtcctagctacctgagaggctaaagcaggaggatcaattgaatccaggcgttctagtgagg
cttcagtgagcaagattgcaccattgcactccagccggggtgacagagtgagaccctgtttcaaacaaacaagcaaaagcaaaaagatccatcaatttgttgctccattcctttacttactatcttgctgtctacagctaaagtcagactg
catgatttggaattccaactccaccatcacctgggtaatgttgagatttattacatagcatctctgtgcctcagtttctccatatataaaatctacctgtaataatgttgtgaaaattaaaagaaataccagttgtaaatcactcagaactgtgcc

tagcatttaataaacactctgtaggtattagcttatttctttataaaatttattcataaaatgaggaagttagctcatgtattgtcataccagctttcactttgaatgagtccataattgcatttttaaacacaaaaagagaaataataggaaaaaat
atgtatacatcaatgttacttcaatcacgggatgcatagtaattaggatttctttgccctaattactgatgtggataagccggtgggccatgtcaatggcatctcaagctgttatcacaccatgtgaaacacacgatctatgtaaaaactatgc
ttggttgcacagcttcaaggaaagcactagtctactcacttccttttcattcagcatgctatgtctaggcccctttcatataaaatatttggagctgccattgtaggtagtgcaggtggtcttgttcaaagtgtgagtatcacctcccactgcca
agactgcagctgctttgattttaactgaggatctagaaaagtgctgggaacacaaatatttgctgacctttgtcacaagtagcacctaatctccttcctaccagacctgctgccttataaaatgatgttctaagtgtataatgacttcctctgct
tctgcaactgtgttttggaacaggagggtgagattagtagccacacacactgaaatagtcatatgacttccaaagtctgtaatttttaagcatatacaaaatagtgaaatttctccatatgtgtataatatgtatatatgtgtgtatacatatatat
agagagagagacagagagagatgggggagagagagagagagagcgagagagaggcagagaaagagatccttattttggctgatatctattctatacttctttgaaatagtgatttttagtaaattggcagtgttgccagatgacgact
tagagttttaaccacctcattgttatatgtaccatggtagctcatcttttctttcaggtgaaagttgtgatactttgtcctcacgctaaacatagcccccatgttgttccaggaaaatagagatcaataaaatacaaagtacttgtttgaaagtga
aggttaagttcaaattcaatgacctacgctgtggtaaaacgatcattcattaatgatattaagtattttctttttaatgtcagaataaaaccatttatttgcagatttaaaaatatatatgtgtatgtaaatctccttggtttatacctcaaatatgaaa
ataggaatgaccctgtctctatgctcttgcattcacttatattctagcttgttaagagagaagtttggggcaaaatgtgctttataaatagtaaaaactgaggtaaactatagcacaaagtgtgctatataaacagtaaaaaccgaggtaaaa
atatggggaggacaaaaaagggctaaaatggtgttggaatataagaaatagcagtaaaagaagctcctcacagcaagacagtacagggaaatgtctggagggacaaagatttacttacaatcaggcctcaaccaatatgagatcga
ctttcttttattgatatatgacgtgtttatatgtatcaggtacatgtgacattttgttacatgcatataatgtataatgatcaagtcaggtatttggggtattcatcacctcaagtatttatcgtttctatgtgttaggaacatttcaagacctttcttctac
tattttgaaatatacaatatattgtttttaactataatcaccctactttgctattgaacatcagaacatattccttctatctaactgcctgtttccattaaccaacttctcttcgtctcccttcctcccataacgtacacacctttctctacctctggtttct
gtcattctactctctacctacatgatattaactttttttagctcccacatatcagtgagaacatgcaaaatttgtctttctatgcctggcttatttcgtttaacataatgacttccagtttcatccatgttgctgcaaatgaaatgattttatatttgtatgg
ttgaataatattcaattgtgtgtatacaccacatttcttcattcatccattgatggacacttaggttgattccatatctttgctaatgtgaatcgtgctgcaataaacagtgtggtaccagtatctgtttattattctgaattcttttcctttggataaat
acccagtaataagattgctggatcttatgatagttcaatttttagctttttggtaaatctacatattgttttccatagtggctatactaatttgcatgcttaccaatagtgcctacgagtgccttttccttcacatccttgccagcatgtgttatttattta
ttaatttatttaaatcatgatgattctaactggcataagatgatatctcattgtggttttgatttgcatttctctgatgattagtaatgttgagcaatttccccatacccattgtgcatttgtatgtcttcttttgaaaaatatcctttgccaatttttttaatg
agattatttgtttttccactgttgagttgtttgagtttcttatatattctggagattaattccttgtcaaatgtatgcaaatattttctcgcattcaacagcttgtctcttcactgtgttgattgtttcctttactgtgcagatgtattttagcttaatatagtcc
catttgtctattttttgcttttgttgcctgtgcttttgaggtcttggccataaaatctttgcccaaagccatgtcttgaagtgtctccccttatgttttcttttagtagttttatagttttaagttttatatttaagtctttaatccatcttgagagtacgttttatat
ggcaataaatagggggttgtattaggccatttttgcatctattgtacaaggggggttgagttcttgatttgattctccatttggttgctgttggtgtatagaagagttactgatttgtgtgcattcatcttgtatctggaaactttgctgaattattttatca
gttctaggagctttctggaggagtctttagggttttaaaggtaaatgatcatatcatcagcaaacagtgacagtttgacttcctctttactgatttgaatgccctttatttctttgttttgtctgattgccctggctaggacttccagtactatgttga
agaggagtggtgagagtgggcatccttgtcttgttccagttctcagagggaatgctttcaacttttccccattcaatattatgttggctgtgggtttgtcacagatggcttttattacactgaggtatgtctcttgtatgctgattttgctgagcat
tttaatcataaatgtatgctggattttgttgaatgctttttctgcatctattgagatgatcatgtaatttttgtttttaaatctgtttatgtggtgtatcacacttattgacttgcagatgtcaaaccaaccctgcatccctggcatggaacctgcttgat
catggtggattatctttttgttatgttgttgaatttggttatctagtattttgttaaggattttagcatctatgttcatcaaggctatcagtctgtagttttcttttttggttatgtccttccctggttttggtgtgatgctggcttcatagaatgaattaggg
caggttccttctttatctgtcttgtagaatagtgtcaaaaagatggggaccaattcctctttgaatgtctagtagaattctgttgtgaatccatctggtcctgaacttttttttggttggtaatttttttaattaccaattaaatctcactgcttgttattgg
tctgttagtaaatttattgaggctcatttcatggcctatcatatggtctaccttggagaaatttccacgcgctgttgaatagaatgtgtgttctgtggttattggatgaaatgttctgtatataactgttaagtccatttgttctaaggtataatttaaa
tccattattttttttgttgactttctgtcttgatgaactgtctagtgctgtcaatgaagtattgaagtcccccactattattgtgttgctgtctatctcatttcttaggtctattagtaattgttctataaatttgggagctccagtgttaggtgcatatatgtt
taagattgtgatattttcctgttggacaaggccttttaccattatataatgtccctctttgtctcttttaactgctgttgctttaaagtttgttttgcctactcacttttggtggccatttgcatgaaatacctttttcacccttttaagtggagcatttagg
ccatttacattcaatgttagtaatgtaaatgcctggctatgccaggcaggaatggcctgctcagggacccagtgagctcccagggcctttctcactgcttcctctacacctgtattttgctcagctctctaagttaactcaactccaggtaag
gtcagaaacttctcccacaaacagaccttcagtttctccagtgggagtgtgtgtttaggagaggaggatctccctttcccacttccgcagttggggcactcacagtatttggggtgtctcctgggtcctgcaggagcaatctacttccttc
agagggtctgtgggtcctctcaggattgctggtttgttcttgcagtctatctgcagctaaaattcacaatgcgagcctccacatgctgctctgtccatctgagttggagctgcaatctaggcctacctcccatccgccaagatcctgagaa
gttggggattatatttcaacatgagatttggagggaacaaacatccaaaccatatcagcagtccagttttattcttttgcacatggatatccaatggatatccaatttttttcagcaccatttcttgaacactgtgtcctttcctcaatgtatattctt
gtcacctttgttgaaaatctgttggctataaatatgtggatgcatttctaggttctctattctcttccattggtctaagtgtctgttctaatgtcagtatcatgctgtttaggttcctatacttttgtaatatactttgaaatcaggtagtatgatacctct
agcattgttctttttttcacaggatgactttggctactggagctatttttttaatttcatacaaattttagattggtttttccatttctgtgaaaaaatgacattgctattttgatagcggtcatatagaatctgtagattgctttgggtagtatggccattttta
acaatattaatttttctgatccatgaacataggatatctctccatttgtttctgtccttttcaattcttttatcagtgttttgtagttttccgaatagaggtttttcggttccttggttaaatttatccctaaatattttatttttttagctattgtaaatgggattg
ccttcttaatttccttctcagctagtttattattggtgtgtagaaacattgcttgtttcatatgctgattttgtctcctgtaactttactgaatttatttttcagcgctaagagtgctttggtggagtcttggtttttctaggtataaaatcatgtcatctgga

aagatggacagtttgacttcatcttttcctatttgcatgccttttctttctttctttctctgacctgaaagctgtggctaggacttccagtattatattgaataggagtgatggaagtgggcatccttattttgtttcagttcttagaggaaaggcttc
cagcttttcccaattcagtatgatgttagctgtgggtttgtcatatatgacttttattatattgagatatgttccttctatgattggctgcagatagtttttatcatgaaggaatgttaaattttatcaaatactttttctgtatctattgagatgatcatat
ggtttttgttctttttgttgatgtgatgtatcacatttattgatttgaactgttcttgcaaccctagtataaatcccacttgatcatggtgtattatcttttggatgtactgttggattcagtttgctagtattttgttgaggattttcacacctatgttccatc
agggatattgacctgcagttttatttttacttgcatccttgtctggttatggaatcaggataatactggcctcagagaatgagttagggagaattccttcctgtttaatttttaaaagaatttaaggagacttgttagttttctttgaaagtttggtaa
aatttgacagtaaagccatctggtcttgaaattttctttttgggagactttttactactgattcaatttcattattcattactggtctgttcaggttttctattctttcttaattcaatcttggtaggtggtatgtgtccagaaatttattaatttcctctaga
tttttcagtttttttagtgtatggttgttcataatagtctctgataatcttttctatttctgtggtatcggttgtaatatattttcttttttatttctgattttctttatttcagtctcttctctttttcttggttagtctagctagagctatatccattttatctttttaaat
aacaacttttcattttgctaatcccttgttgttgtttttctctatttcattaagttcttctgttcttcatttttttctttccttctgctagttttgggttttgttcttgcttttctagcttgttgaggtacatgattagattatatatttgaaatctttctatgtttttaattt
aagtgtttattgataagcttctctcttagcactgcttttgctgtatcccatggcttttcctatgttgtatttcaattttcatttgtttcaaaaaatttttgatttccttcttaattttcttcttgacataacgttcattcaggagcatgttgtttaatttccatgtat
ttgtataggttttaaagtttttcttgttattgatttctagttttattccattgtggtctgagaagatgcttaatatgatttcaaatatttcaaattagttgagacttcttttgtgttctaacatatggtgtatcctggggaatgcgtcatgtgctgatgagat
aaatatattctgtagctgttggatgaaatattgtgtaaatgttagatccatttggtctaaagtcaagtttaaatccaatgtttctttgttaattttctgtctagatgatctatctgatgctgagattagaatgttgaaatccccaactattattgtattgg
attctatttcaccctttatatctaataacatttgctttacatatctggatgctctggtgttgagtgtgaatgtttagaattgttatatactcttgctgaattgttccttttatcattatgtaatgacttttgtctcttagtgtttttttatttaaagtcaattttatct
gttataggtataactacttttgctcatttttggtttctctttgtatggaagatctttcatctttttactttcagtctatatttatctttacaagtgagataagtttcttataggcagcttacaactgggtcatatttttgttttaaatctatttagccagtctatat
ctttcaagtaacaatttaatccatttacatgcaaggctattattgacatgtgagagcttattcctgccatttgttaattgattcctggttgttttgtttttttttttagtttgtttgttttgtgtttttttcttccttccttcctttcttttcttcattctctctctctct
ctctttctttctttctttctttcttattatttatgattgtggtttggtgattttctgtagtggtaacgtttgagttctttctcttctttatttgtgttttgctctaccagttagttgtatacctttgtgtgtcttcataatggcagtgtcctgtcactttcatgtgtag
gactcccttaaatatttcttgtaaagccagtctagtgatagtgaattccttcaaatttttcttgcccagaaaagattttatttatccttcatttttgaagaacaatgttgctgggtataacatccttggtgtcatttttttctttcactattttgaatatatca
tcccatcttctcctggcctgtgaggtttctgctgagaaatctgctgttaatctgttaatcttgtggaagttcccttataagtgactagacattttgcttgctgttttagagttcttactctttggcttttgacagtttgactataatgtgctgtggagaa
gagctttttgcattatattatttgggaatttctgattttcatttgtctggatgtctacatttcttgctagacttgacaagttttcagctattgtgttgttacgtaggtttcctacacctttcatgttctctttaccttctgggaaactgcaaatttgaatatttg
ttcactttatagtgcctcatatgtcacatacactttgttcattcttttttttcctttctatcatttgaccagctaattttaaaagacctgtcttcaagttctgaaatttttcttctgcttgatctgatctattgttgaagtttttgaatgtattttgtatttctttca
ataaactcttcagttacagaatttctgtttggttcattcttgttgtatctatctcttgggtaaatttctcattcatatcctgaattgtttttctgatttctttttattgtttatctgagttctcttgtgtctcactgagcttccttaatatcaacattttggattctttt
ttccagaattttataaatgtcttttttttccactggcttatgttgctgaagaattattaaattcctttggaagtgtcatatttcttcacctttttatgtttattttgttcttgcattgatatctctgcatctggtgtaatgattgctgcttccagttttgtaaatttg
tttttgtaggggaagactttttctgaagatgtatcttggtgttgattgagtagggcactttgattttgatttctaggtgcatacagcagcatagtcttcatattattttttttcagctgccaacagtatcatcagtgtctgtgatttttcagtggctta
gggtgcatttgttagtgaaggtggtgatgaggttgtgctggagacaagaatatcaagtgggccagtccttggccccagtggtggtggtgggcctagcattcctgtccttgggccccatggctgtgtacattggcactgatggtaatgag
tctagttgggttgattctttggccttcaggcagtttgcttgggtgtcagtagtggccatagtagactgggctggtgtgtgggtcctcaggcttctggacagtgggtgtggtgtgggtgatggctgtagcaactgcaggacaaccctctgg
ctcccaagcagtctgcactggtgttggtggtggctgcaagggctgaatgagccagtcttcagacccacaggtggcacatgcaggaggtgccagttgtgatggcagtggcagattgggtaggcccatcatcaggcttctgggaggg
ttgcttaggtaccagtcatggtggatggcacaaggctgtctctgggttcccagacaacttgcttaggcattgaggaaggcaaagccaggctgggtgttcctgacctcaggccactcagtggtgtgtgtggacaccggctatcataggc
tggagtggagtgatcctcaggcccccagcagaatgcctgagtgaattagcagctactgcattgttgccctgctcttgaaggggtgcggggtgtgttgctttcagtggcagcagctataaatgggttgctggtgggcatgtgtttcagcc
ccagatggtggctgcaagtggggtagcattttcttagggagcttataaatgtgtagcagccctgctactgggggtgggggagtccacttctaatgacacactttatccctggtgacagcagccagcagtagcactcagctgtgggtgg
gggatgtcagcagagcttcagggatgtagagatgcagatgatgttggggtctgaggaacgatgtcatctggtgggtgttgggctctcaaaatggcaccttgctgctcagtaatggggatggcaggggtgggacccagtgagagctc
cttctctggagcaatgcctttgtgcgattgctgggcagttccttatgatagtctcaaggcctgcatggtcaagggactctcctgtggctaggattcctggagtccataatgcagatgtggaccactgggagtcagtcacttaactttcgatg
cactggggagcctctctggcctcacagaggatccctactgggtaggcttccttgcttccctccccttcattgccttagatgcttcctgtcacttctctgttgaattccagaattctctcttatgtgggattttctacttactattttggttctttgtgg
aggaggtgagtgtcagctgcccttagtcagtaatcttcaagccccttcccttgaaattttttaagctttatgtttttatttctaatagactgaaagagcataatcatattttaaacactctaaggctatatgggggactatctgaaggcctcagat
gagtaaaagaagtagcgaattagttaactatgtttttcttttcatttcatttcccaaatcaaatcatttgtttcttcaaataaagaatctctatttccccccttatgttcttcagcaaataaaaatgtttgcatctaccaaagtgtatatgaatttgaaata
cgtaatctaacaaacaaggatgttttaagatctctgagtaagatgcatagttgcaaataaaaggagttttcaaggataaaaaatgtaaataaaacatctaataaatgtgtaccatcagaaacttatggtatattctcctggaaaagatgaag
atgaagaaaataagtctacagaagaggagaaattctttggaaattttgcatttttgtggatagtggttttgaagaaagcttagatatatttaaatatgttgctaagatgtgactcccacactttaaagaaaagttcttcattttagttgcccttgca
aattggctaagtttatttttcctcaacaaataattttgatggtttctaaaattgtccagaatttgaggaaatacttagatgtagtagatgacatcaatttatgttaacttagatttattgttggatggttaagttaatcatgtctgttcttgtccagtctag

aagtttctgcttcctcctcatgtgttctattctgtcctcatcacataacatctgaactgtgatttgacagtgtccttattatcagactctgctcaatgacagtttaaacttcttttgatacagtaaaaaacttttccttaaagatgaaattcttaagga
caattagcttgtgtagactcatatgaaaattgcttgcattttggagttacattagctcttctcctttataaattaatgaatcctattggtaccatataatagcatcaacatttgccttttgtgtctcgcctttcaaagggcaattttccaaaaagtatat
tttgcatggctcaggctaaaagtaatgtctccagatgaagttctactttgtgaccattgattgtgactattgataagatttactatgtaaagtataagttactggaataatacatttaagaaatatagtagctattaaaacctgtagtctgatgaaa
tggaatatgttttgaaagacagcagtgaaaaatgagaagtaatatataagtaataaaattgctttttgcaattgagttgaagattgatgccaggatcttattacctccaggatttcacaaaggatttcttgagtcatccccttgctctagctcat
gaacaaagtattgctcaagtaatcgacagattatttgtgtattctcacatgtgattattatttagtgctgtcacgggaaatatcctgcctaagtttactggaggtcttcctccccttacccctccctatgtcctcctacctggttatccttttcaatc
ctattactctttccttgcacactctactccagcttagggcctgggctctgaccattacctctgcctaaaaattcctcctcatgacatccatttgcccaatttcttcacccttttaattaccccattaagtgatattttgcttgctcaatttttgccttttc
aatcagatctaatttattttatatttttaaaatttatttatatatatttagaaggtatgtactagttcgttatcatgctgcttataaagacatacctgagactgggtaatttataaggaaaaaacgtttaattgtctcacagttccacagggctgggga
ggcctcgggaaacttacaatcatggcagaagaagaaacaaacatgtccttcttcacatggcagcagcaagtagaagtgccaagcaaaaggggggaaaagccgcttataaaaccatcagatcttgtaagaactcactatcacgagaa
gagcatgagagtaacctcctctatgattcaattacctccctccaaatccctcccacaacacgtgaggattatgggaactacaattcaagatgagatttgggtggggacacagtcaagctgtatcagggtacaagctcgggttccttacat
atgtatgttgcatagtggtgaagtctgagcttttagtggaccatcacctgaaaagtgaatgttgtatccaatccgtagtttttcagtcctcaccatactcctatcctcccacctttctctagtctcaagtgtctattatttcattctatatgtctatgct
tactcatcgtttagctcccacttggctctcactaaatgagaacatactgtgtttgactttctgtttctgagttatttcactgaggataatgtcctccagttctaaccatgtttctagagaagacatgatttcattctttttatggctgagtaatatatat
gtatataaatatatatacaccacattttctttatccaatcgtccattggtagacatttaggttgattccatatctttgctattgggaacagtgctactttgaacatactaatgcagatatctttttaatatgatctctttcactttgggtccatacccag
tagtgggattgcaggattgaatggtagttctgtttctagttctttaagaaatcttcatgctgtcttccataaagtttgtactaatttacattcttaccaacagtgtataagcattttcttgtctccacatcctcaccaacatctgctgtttttgaactttat
aatgatagccattctgactggcataaggaggtatcatatggtggttttaatttgcattctctgattattagtgatgttgaacttttttttgacgtttattggccacttgtatgtcttcttttggaaaatgtctgctcaggacctgagctcactgtatttgt
cagggttccctagagagacagaactaataggagatagatagatagatagatagatagatagatagatagatagatagatatagatttagatataaaagggagtttattaagtagtgttaacttacacaatcacaaggtcccacaataggccatc
tgcaagcagaggagcaaggaagccagtccaagtcccaaagctgaaggacttggagtccaatgtttgagggcaggaagcatccagcatgggagaaagatgtaggcttggaggctaagccagtccaggattttcacatgttttttttct
gcctgctttatatcctggccacactggcaactgattagatggtgcccacccagattaagggtgggtctgcctttcccagcccactgactaaaatgttaatctcctttggcaataccctcacagacacactcaggatcaatactttgcacact
tcaatccaatcaagttgacactcagtattaaccatcacaagtccatcgcttgtcaacttgagcctatacacatcccctgaggtcatacattatcttcaaataaagacaacaataagtcataattatgcctaacatactaaaactatccttcgtg
caattgaaaatgcaccaatccccattgcaaatgctattacataaacttaacaatacttaaatgctgatatgaagtcaataaatcttatgtcacatgataaaggaaaaaggaagtaaaatgtatatattttcttagtacaagtgtatacagacac
gaaggtgtttttaacaaaagaaggaggaaatactcatgacaattacagtccttgtatctgcaattggtcacgtgatcgtagctggtattgatgactaccttcttctgctacatattctgtattccctttgccctcagcaagcacctcagcaggt
cttgtcttttttttttttttttcctggtggagtgacccaaacctttatttgtagtcctgcctggaatggcctgtagtagtttcccactgaccttaattccaggacatggtaaaactaagagacaccctaatggatctcctatattccatgcatacgc
ttccttacctccattgtggagttgtagactgatttcatcttgatagtctgggttaatcaccgcagccaacactgtaactctcttcttagcctgttgacttaaaggtgggaggagcccaaagtgtccaggtggcaatcttaatttgcagtttaatg
gaatcattattgtgtcacctggcggcagtgttcctccctctggaactaagacctctaagccagcagaacataatgccatgggaacaggaagcaaaaattttgctagtggaacactaggggtgatggtgagtggtgccacttccacttcc
acttccaccctttgattcctggacccgtgaatcctgtctatgggagaaacagtaccatatattggacactgattcagagcatacacagccttctggagaactttgccccaggcctgcaaagtattgtcacctagttggcattgtaatcgtga
cttcaaaaggccattcaaaagctgcttcaggatgatagggatcatggtaagaccagtaaattccatgagcatgagcccactgtcacacttctttagccataaagtgagtggcttggtcagaagcaatgctgtgtggaataccatgacag
tggatacggcattctgtgattccatggatggtagtcttggcataagcattgagttcaggattgggaaacccatatccagagtaagtgtctattccagtgaggacaagcctctgccctttccatgatggaagaggtccaatataatcaacct
gccaccaggtagctggctgatcaccccgaggaatggtgccatatcaagggcttagtgttggtctctgatgctggcaagcagtggctgtggccaggtcagctttggtgagtggaagtccatgttgctgagcccaagcataacttccatc
cctgcaactgtggccactttgttcatgggtccattggacgatgacatgagtggcctgggaaaaggctgagtggtgtccacagaattgatcatcctatccacttgattattaaaatcctcttctgctgaggtcacccattggtgagcactcac
atgggatgaaaatatctttgtagtttttgactaatcatagaggtccatccacatacctcttccccaaatttctttgtcatcaattttccagtcatgctttttccaagtccctgaccatctagccaaacctttggctacagcccatgaataagtatata
atcgcaaatctggccatttctccttctgtgcaaagtgcacaaccaggcacactgctcaaagttctgccaactaggaagatttcccttcaccgttgtccttcagggatgtcctagaaaggggctgtagtgctgcggctgtccacttttgggt
ggtgcctgcatatcgtgcagaactatctgtgaaccaggacctagtcttctcttcctctgtcagctgatcatagggaactccccatgagaccatcggtgcagtctcagggagagaaagcagtgtgtcaggagtggagaccatgggcatt
tgagccccttttctcatgtaacttactagtgccttcaggatctgctcaagcccaatcgcatatataccacttccatttgctgatgcaatgctgctgtgcacagcccactttatggctagatggatcagaaagcacccagttcatgataggcag
ttcaggtcacatggtgacttgatgactcatagtcaaatatttcgtttccaccaaagcccagtaacaggccaagagctttctccctaaaagagagtagttatctgcagaagatggcagggccttgctccaaaatcctagaggcctccgctg
tgattcagcctgccaaatgctccaaacagcatccctatctggcactgacacctcaagcatcattggatctgctaggtcatatggcccaagtggcatagcagcttgcacagcagcctggacctgtggcagagccttctgttctggactcc
actcaaaactggcagccttttggatcactcgataaatgggccagagtaacacacccaaatcaggaatatgttgcctccaaaatccaaataggcccactaggcgttgtgcctatttcttggttgtaagagtggccaaatgcagcaaattat

ccttcactttagaaggaatatctcaaagagccccaaaccattggacccctataaattttactgaggtagaaggtccctgaatttaatcagatttatttcccatcctctgtcccacaaatgtctcatcaataaatccagtgtgtttgctacttcttg
ctcactagttccagtcagcatgtcatcaatgtaatggaccactgtgatatcttgtggaagtgaaaggtgagcaagttctctccaaataagattatggcacaaagctggagagttgatatactcctgagttcatcattttctttcatcactttgtc
cagcgaacttacaagcaaccaaccaacttcactatgttccttgattctccacatatagtcaaaagtattatgtatacagtagctaaattccttgcctctcacaagcgttgaatcaagagtgtcaaatgcatttattttgcataactctctaaaca
attgatgccaaggattatcagtgttctccataccattagaagaagagtccttagcatatttgggtctaatcatattaagcagccaactccagaaaccccaaaaccaatgaaagaactccatccttaatattctgttcctctagaaccactcct
ggtaccaaaatctgtattagggttctacagagaaacagaattaataggagatatatatctatatctatatctatatctatatctatatctatatctatctctatatctatctctatatctatatctatatgagagtttattatgtagtattgactcacatgt
tcacaaggtcccaaaacaggccatctgccagctgaggaacaaggaagctagtcggagtcccaaagctgaagaacctggagttcaatgttccagggcaggaagaatccagcatgggagaaaaatgtaggctgggatgccaagcc
aatctagccttttcacatttttctgcctgctttatatcctggctgtgctggcaactgattagatggtgcccacccagattaagggtgggtctgccttccccagcccactgactgaaatattaatctcctttggtaacaccctcacagacacacc
caagatcattactttgcatccttcagtccaatcaagttgacactcagtattaaccatcacagtcactttctaatgaggttggtttttttgttgttgagttccttgtagattctgtatattagccttttgttgcattcatagtttgcaatttttttttctcattct
gtaggttatccgtttactctgttgattcggtcttttgcggtgcagaagcattttagtttaattaagtcccatttgtctgtttttgtttttgttgtgtttagttttgaggacttggtcataaattctttgcctaggccacatctaatttttaatcctcttttagag
cctgtctttccctccaaactaaatattattttacccccatttgacacatgagctaattgagccttctgtctgttaagtaagttgactagttcagtgtaagttgtttctaatttttttgccatcataatgctttagtaaacatccttttcatactctccttaatt
ttcccccaggatggttttctagatgtggacttgctggatccaaagtgtgcatattttccaaatgtatccatatattgtggctttgtgttttaatagaggtagcagcaatattagaggagagaattaaggaacgattctagtgtagcaggacaag
ctgcatacaaaacccctcagacaccgagttaaagaaggaagggctttattcggctgggagcttgggcaagactcacatctccaacaaccaagctctcagagtgagcaattcctgtcccttttaagggcttacaactctaagggggtac
atgactgggggctgcatgcactggtcatcagaatggaacagaacaggacagggattttcacactgcttttccatacagtgtctggaatctatagataacataactggccaggtcaggggtctatctttaaccaggcccaggatgtggtc
tcaggctgtttgcctgtggatttcatttctgccttttagtttttacttcttctttctttggaggcagaaattgggcataaaacaatatgaagggtggtctcctcccttactaggaagagagaaataccatgtgagaatgaaatacgtatttgtgtta
ttggatttgacatcggggttggaccaagcatcttctatgaacatctataagcacgttctccttatcaaaaggtactttgcttggtgctttaaaagataaatgactatgtagcttaggagatttcttcataattgaaaaccacagcagctgttattt
gaacatacctgagccccttctggaccactgtaggagaaggaggaagagaggtgaaagtgatttctgctgctgagggcatgtatgatgacaatgggtatttgtcctaactttgacactacaaagtgactggagctttaggcaagtagtta
ctccaattgtcttttcagactctttctttttgaggtatttgacctctttgcggcatcagcctgctgctatttcacatgcacctagaagggactttaaaacatgcacagaattgaacaaagcagtaaccattcgttggtgtcccttaacaatatgtc
aaatactattatttttgaaaggtaaataaatccttgggcttagtggatcctaaatgctaatttaaaaatctgaactcatttccatcccttattactttaaacaaattgctaacctgaatgataatagaaactttttttccttttaacag*TTGGAGG*
CATCTATACTGTGATTCAGACAAAGGCCAAAACAACAGCAGATGAATGGGGAGAGAACTATTTTCTGATAGGTCCATATTTTGAGCATAATATGAAGAC
TCAGGTGGAACAGTGTGAACCTGTAAATGATGCTGTCAGAAGAGCAGTGGACGCAATGAATAAGCATGGCTGCCAG **(SEQ ID NO: 1803)**

Exon 1-2 of *HGD*

AGTGAAGCGCAGTGAAGCAGTGGGAACCGGAATATCCAAAGAGTGGTTTGAAGGAGAAAGAAGCATTGTGGCTTTATATCCTCTG
GGCCTGGGTTTCCTGAAGTCACCACACATAGAGGAGAGAGAAAATGGCTGAGTTAAAGgtaagaaaccatctgacaagtttgctatggcttcttcagccaaaatttat
ggaagtttctgagaacttcagggtatcagagttggaaaagacttagagaggccctggttcaaacccctgtgttcagaggcttccctgtggtgtctgtgggcacagctggagctgaagaacaagggaaaggtgataaaggaaagaaa
tgaagctcttattggtcatcccttgcctttgttcatttcaattattgtcttggaaatctgtagacatcctaagagctcccctccactacggacatgcacgtttctctctatgtctttttcccaacccctgctctgacaccttcccattgctgcagcct
catggaccacttccggtgaccagttagatatgcttgatctatttagttaattaaggctctgaatcagcccttggccagaagaagcacctcctctaggaaatcattccttctttttagtaatacgcaattgaaaagtatgtgtgggttatgtgctg
ttaggtgctaagcattttagctatgaattaactcataatcctcaaaataaatctggggtatagagagaggttcttgtttcaattagaaaagtaatgcttggagagactatgatttgcccaaggtcactgccaatagatgacttgattagaagca
gggtatgtttgaatgcgaagtccaggctcaaaaaggggccttagtggctcctactagactggtatcttaaggcaaggaaaacagaatatgtccctaatatctaaaaccatgcctagtgcccaggccaggcttgaatcaatgctcaacg
catgcatgtgtaactgaacagaactgcccagaaccttggctcaggccaagtaggccagaccaccccacaacagacatcttccctgaatgcagggtttgagaagtgccctaatgccctctaactgatgctgagtttgaagaatccatca
gttaaggggaggtgtttctgagtttgaagaatcaaacaccttcctaattggtaggaatcctgaaaacactccacaccggagtaaaagactttggtttttatctttatgcctgactaaaagttaactctcattctttcattcagcaaatattttgtg
ctagtgcatgacactgcagattaatatcgtagatgaactgagtctgggctaggtagagtcaaaaaaaaatttcctaaaggagcctgtacctaaactgagtcttcagagagaagttggaactggcaaagtgaggaaggacggggtctg
gagggaagagtgagtgcaaaggcacagcacatacagaatgcctgcatgctctagaatccaaaactgtgtcagccaaatttgagttgatggagggacagccaaactgagatacccaggcggtggttggatttatggatctggatttca
ggggagtgggcagggctaggggaatagaactgagggtcatgatgtcagtggtaggagttgaagtcttggaagaaagagtagatagaaactttgttttctgatgaacattctatgaggaagttcagcagaggttggaaacatctttacc
atttataaataaattttatgtaaaatcctatcctgctgtcatgtttaagcaacttattatgcacagcaccaggacatcaatgtaaattgcatgccagtctttgctctcagtttcaaaggcattcttccctgcctgctgtctaatctttacaaagaaag
gttaactccagaggtcaggaatctgaacacagaaaaccagattgcatcatattcattctaatcaataaacattatataatttttaaaactgtttctcatcttttatttcttcattttaatgaatatgtcaacacaaagccgtaaatgaaatcacattta

cacaaaagtacaactgggtaagaaaattcaaaagataaagtaataaaatttaaaatgaaatacctatgtatttgagtaagttgcttagcttctcccaacctggtttcctcatttctaaaacgaagaagaactggcttaatattcaggctc
tgggccccaccaccaaatccttgattcactggctctggctgtatgtaacaagacttctggagtgatttctaatgattagccagagacaggacctactattctactgcgtaactgatttcctctgctgctgtaaaaaattaccacaaatttattggc
ttaaaacaacacaaatgtattatcttgaagttgtggaggtcagaagtctgaaatcggtttcactgggctaaagtcgaggtgtcagtggctgtttcttctggagactctatgagagattgtgtttcctgcctttccatggattctgtttcctct
tctagaggctacccacattccttggtttgtagcctcttcctcgatcttcaaaaaaatgtatctctgctttgtaatcacgtcacatttctccttgcctccctcttataatgacacttgtgattacatttaggatacac
ccaaataattaggatatctccccatccaagatccttaattatatatgcaaaagtctgtttggctagaaaagataatactcacaggtttcaggtttaagccatacatatctttcagagtgagccattagccattcagcctactgcgataa
tgtgcaagttccctttacaatctatgttctatgttccatcaaggttaattagaaatattagactctttttccacagaaaaaaaaaaatgtagcagtagtttcatctctacctgacttctctctcttcctcatgcccagtaagaggaagtgaggt
gcaccatctgttcatctcatcaggtgttacctagaatcagagtcaggagaaagcacctcataaaaactactctgacaagtaaacccatatcctgggggaccaggtgagaattaaaaagctgtatactcataagctgccaaagagctgatggt
actgaagtcacatggagctcaggaatttattcttgcatttagaaagtaatgcatatgttttctgtcggatcaaattattcagaaatccgaactgtagaaaaacatacaaacacacttgcttaatgttag
cttttaatctattatttttgtacacagtatactttacataggcactgaaccagctggaaatgaatgcttggatttcgcagctgataaaagcagagttcagattctgtcctgagctcaatgtctggagaaaagcagctgtgaatactgcttcctta
attcagatgcctttggccgtgaggggcatggcatttaaagctgcaataaactctgtactttttctctcaggggattagaaaaaccaattgatttgatagtgactcagcagtttcacctcctctttagccaatgccctgcagctaagattga
gaaacggattcagaactcacacctcttggtgctggctccctccaaacccatcagctgttcttctacttcttcctatttttaaagcacagtctgaatcaagtgcttcagaaccatttatactcatgttccaaacacataacccctg
aaccctcttctaaggccaccagttacaaagaatgggggaaggttaatgagtaactccaccagtgcctgccagaaagctcagctgtactccagccttcctcctctctgcagagagaaacaatgcggcagttgaacaaagcgaaggca
ctgtaatgcccagtggcttcagtcacaggctggtctaaaattccatttcacagaattgaagataatttgggggttattaccaattctcagcagctcaaattcaatccgttgatggtgagcatgggttcatttgagagtagaaaaattggatg
cagtggagaatcagaatgttggatggtctctgccatttcccaaaggatatcatcctaactgcatgctgctttaacaaatctgcatgctgctatacaaatttcaaatatacaaaaatattgagtctgaaggcaaaaa
gaataatttttcacttccattgcttcctaggtcataacattcacatctgatatttcttccaaatacagactgttgttggcctgatcatcactcacacatccaggcttgttacatctccacgccactgagtatcagggaacctgcccaatattcac
gtaggttctttctatttttccctaagcgtcggccaacttagaaataaagggacagagtacacaaaagagagaaatttaaagccgggcatccacattccggtaggttccgtgatgccccacaagccacaaaaaccag
caagtttgtattagggatttcaaatgggaggcagtgtgcaaataggtgtgggtcacagacatcaagtactttacaaggtaatagaatatcacaaggcaagtggaggcagggtgagatcacaggaccacaggaccgaggcgaaat
taaaattgctaatgaagtttcgggcaccattgtcattgataacatctatcaggagacagggttttaggatcaactggtctgaccaaaatttattaggcgggaatttcctctccttctaataagcctgggagcgctgtgggagactgggtct
atttcaccctgcagtctcgaccataagagacaggcgcacctggaggggggctgtttataagcctatacctccctggctcgtattctcttctcaggagatgttccatgctgaggaaaaagaattcagcgatatttctcccatttgctttgaaag
aagagaaatatggctctgttctgcctggctcaccagcagtcagagtttaaggttatctctcttattcccctgaacaattgctgttatcctgttctttttcaaggtgtccacatttcagttgctcaaacacacatcagctgtacaatttgtgcagttaat
gcaattattacagggtcctgaggcaatatacatcctcctcagctgacaggattaaagagattaaagtaaagacaggcataaatcacaaaggatattgactgggggaagtgataagtgtccatgaaatctttacaatttatgtttagagattgca
gtaaagacaggcataagaaattacaaaagtattaatttgggggaactaataaatgtccatgaaatcttcacaatccacgaacttctgccatggcttcagctgatccctccgcttgggagtcccagactttccgccaacaactgagaactctgg
agggcctacccttcacccacttccaccttctttatatctcttaaggagctgttgttcagatggttacatttcgcttgctgagcaagactgcatttcattatctcctgggctaaaatcaattggtatttaagttagagcaaaaaggttctaaccctat
ttacagcagcacacatacacttacattagaattggcttaagggaaaaattcaatcaatgccacaaatttccactgattacctttctatgagccatagtctcagcttcatgcagcttatagggctagcaaaaaagccagacagt
gcacaagaactcacaagtgttgagtgttgcaagagtaagtttccagggctctccatgacttacctgggttgtaaagtgaagagatggcaatcagagtaggttccattgttgtcattgtgaa
aagatgctgggctggcgtgtctcctgaaggccacccatatctttaaacttttattttcatgcattcgcacacaaatatagatgtgttccatagatgaggcagatagcagaggcttctgcctacagcaaggctagtttggagaggggtttgctgtc
agggttggacccacacacctcttcatagatgctttttctttttttcccagagtttgaaagggaaatggagtctcaagttcagggaggttacagtctaaggtggagaagtctatgcaatatcc
agcactcttctgattaattagaagtctttaaaattcacagaaaccatattgtcttccctaaatgattcatgtcctcaagacacagtaaagatattatatgtggaacatcatacacaaactaggtccatatagaaattcttacctttagtaaactgcc
aatttctgaaatgctctttaaatcctgctaatggtggcatttgtgtctttgtcacctacag*TACATTTCTGGAATTTGGGAATGAGTGTTCTTCAGAGGATCCTCGCTGCCCAGGTTCCCTG*
CCAGAAGGACAG (**SEQ ID NO: 1804**)

Exon 1-2 of Lp(a)

GTAAGTCAACAATGTCCTGGGATTGGGACACACTTTCTGGGCCAGTCCCAAAATGGAAGTGGTTCTTC
TACTTCTTTTATTTCTGAAATCAGgtaagacatagttttttaaattataagaattatttttctcccacaatgtagtaaaaatacatatgccattggctttatgtgcaattcatttaatttttgattcatgaaattcccagtt
caaaatcttgtatatgattgaaaaaattcttaaaaaaaataagtttaatttccccgtgaagactgtcacggtgctggaatgaatgggcagaaaaaaataatggttgatttttctaatctaaaagagtgtgcctacatgatggccagtctggctgaa
aaataaatagccattgtagctaactatgcaaaggatggctaagctcttcgcttgttccattaatttatatcatctctgttcaggtgccatgctcccctcactagcaagttgaaacaatgaaataactcttttgaatatgtttggttcctt

gacctgttcatggagtgggactcagcatttctctctttgttatggcctgagtaaggctttccatcggtatacatttgcttcttatccctggagaaattatacacatccatttgccagatgatatacgcatataatgattcaacaaatactcagggt
atttgttgagtgggttaggtccccacatttttatacatacatacacacatacacaccgtgtgtgattgtgaatgtaagtgtgtgtcctttacaaatactagcttatttagctcatggtataggtagggtagcatagtcatccccattttataaacaa
agaaatctagacttaggaaaatcatgttatttgtctcgtgaccaaattcccaaatcaaggaaataaagaaacctggatttaagccagatttccaagaaaaaatctagggctcttctcactttttcatctttgttccaacatttgaaaaaataaat
ctaaacacattccaatgtaactgaagagcaggttaattgtttgccacttgcagaatccaattaagaagagagaagtctggtataaagaaagtgatttgcttccaaagctagcttaggggaagaaatgcagcagtcctgccgtactacttc
actttaggagcagaaagtggcacttttaaaaggcaacagaggaggcgagcaaggattcaggggtccatgctagcttgggcaccttatccaccaggtagttgagcagttgcctgctggtgcctttgtgagcagggtgttgtcccttgag
gcaaatctctggagggtgagagttttgtagtgggcatgctttggtttataaatcacctgtgaactcaggagttccatcttgaagcacatacatagttagatgaacttgccctgcagggagagtctgatgaaagggaggtagatgcttgca
atttaatctataaattaccagataaaattttacaagttgactttaaagtcaaacacatttgaatttagtggaagccattcaagaaaatatcaaagaaaatacagagcaggagaagattaagcaaagagtttttggggaaattggtgtctatgt
ctgtgtgtgtagggagtgcaggggatatgaatattctatttcagcccatggaaactaggatgtagatcactgtgaacttattcagcaggctacacccaaaggctagaacaaacttctctgccacaggattaacatatgttttaatcgacct
ggggggcacattctctgataagctcttttggaaagccaggctttctgtggacgtgttatctttccaatgtgtgctggaatgcccggggagaggaaaaagtttcttttacagccatgctcagtgagaagcggagaaacatcttctattcaca
aattgctaagtcttttacacatgcaaatatgcatacacattcacacaccacagtgaggaagaaattctcacaccattaataaaatacatttacttcagtagcaatatacatctacattttgcctataatataaaagtatttttcctattaaaagattt
gtttaatgtttcttcaccaacaaataaaccctattaaatccccattgccatatgagccctggaggtgaatcagagaaacaaaaggattgtggaaaaatcatcaggttaaaaaaagaaaaattgattctgttttgggatatttcctagcaacat
gagctggggaggggatctcagcagtgatgctctatgaagcataataaaatgacacagttacaggtaacttagttaaagggggaaataaatggaagtttcctctttttgaatatcaattgtagcctgctctgctacatttcaaaaacactctt
caaaatgtttaactgaactcactgtaggaagcaccttattaatttattgtgtgttttgaagtcacactgtgagctatagaatttacccaagcacaactcttcctggaaaagagagttcaaatgagaaacagtgcggggtgaagacatggata
tgggcctaaaatatctatttctcaatgatattttgatatatctatcaagtgctttttagtggattaggttcagaatgcatcagccaatgcctgttcaataatccagttttccagcatagagcatattaaattgaggaaggacaaagtcacagagg
tggggagcaggtggactgtggccaaggactttgcatgaaacagtgagcgtgcatcctcctccttgccctgccctcatggtctgtgtactctcaggaggtcaggacaggcctttctgagaatgagaatctgttcatctgcctttctactgg
atacttgtcatcggcatacaaacacatgttctctgcagtgtgtcatctttcagaacctcccctgaccctgtattccctagaagtctcgctgctttcagagccaggcttctctcctgctgccacccccactgctcttctagtcactctttaaccca
ctccatctgcatgtggcccccaccacacccctcaaagtggtcaaggttgtcctgttgcttaattccatggaagcttggctatcttcattttattagcctcttttggcctctcaccctgtgaaaatcactacattttgtgccagagatggagctgg
catctccaggcttggaagagggctgctgaagctcagccaggtgtcctaaggagcctcaggacaggggatgctcagtagccttgcaatgggaacacagctgagccccacttggccacccctttgccacaaccaggcagaaagcag
cttttgaacagatttgttgcctcagatttgatctcaaagaaaaatcgtgggcagtattggtcccaggttctgctttttacaatttcctctgaaatctggatgcctatcaacaccttggaaaaactgaattctccccaactaatagtggtgtgtca
ctgtagtaagcctagtacaaaaatggccttctttgtggaggagcttcatatcctccatttttttttgcttaatttttgcccaagatgagaacataatttagttcactttttatttattcccaacatcatccatgcaccaacatttttgtaactaaagga
gggaccattcagaagatgcttatcaactgtcaaagtgacagtgttacaaccaatgcacatattgtaagaaatcaaacaatggcctccaaggttcatttctacacagggattagcagatcaacatcaatcttggcaacacagttgccactg
atggtgtcttatttttttatcatgacatggcaatcaagagcaaacatgatttattcttatttaagattttatggttagactaggcagatagctagatatgagcaggaggtggaagcccctgagagaatggaggtctggagaatctgaaaccc
cagagattacccaagtcctgcatgctagacatgagtggaggaggggggaatacctaggtagaaaagaatgccccttaagatgcccagcagtcgctcactgtgcagttaacttttcagaatgctgctagatacatgctgatagggaggg
aagagggcaaaggagaaattcctaagagatacacggttgcagttagtatacatctgagtgctatacaaccttctttgggtggtggcaagaagcaatgcagccattacgtagaattcatatcaaacacctgtatcacaggtgttaaagaa
acaagaaacattgtacttcttgtattcttaataatgatttgcaatattgtctttagtatcactgcaaacctctataaatatgatttttaaaaagtatttctttaggttggaattacttctacgcattgacttatcttcctgggtttcattagccgtacccgt
tgtactttcttccttaccactgtttatctcaaactcttgagattaaagtatgggctcaggagggagcgaggagcttcaggactctcacggacctccagcacagtgtagctgccttatggaaaagtggccacactgttttctgcactggtccc
tgcccctactattcctcactgggcagagcacagccaccctggccctgcctgaacattttagtcagtgttggctctgtgcttctctggggaggaaatccaagagacaacccacagcccctctgccatttcagctgcagcagtaccaccgt
taatgcccttgggcttgagaaagaagggacctggccacttccctgacacctccagcacacagcagggaaagaattccagtttctctttcttgtgagctttcacctgctactcttcaccaggcaaggctcctggcttgggcccacagtgc
aggcacctcgaactcagttgaacatttccactggctgcactctgtgtttttgtggggtgaagctcccagaggtgactgaaagtccttctgccactaacactgcagtcatactgcccttgctgtacttggactagggaaggaaaaaagatc
ctgagtgctttactcacaccccagtgtgccccagccaccctatggaaaagaggccagtgtgtcatccctgcaagcaccctgaggcccctgcccctgctgcccccaagctgtagagccagaatataaagctggcagaaaaatgtaaa
aaggctagactggcttagcctcccagcctacatctttctcctgtgctggatccttcctgctcttgaacatcggactccaagttcttcagctgtgggacttggactgtcttccttgctcctcagattgcaggtggcctattatgggaccttgtaa
tcttgtgagttaataccacttaataagctccccttttgtgtgagtatatctatatctatagatagatataggtatactcactatatatacacatatatacatatactctctctctctctctcatatatatatatatataatctcctattagttctgtccctc
tagagaaccccgactaatacagattttcataccagaagtggttcttgaggaacagaatattaaggatggaattctttcattggttttgggacttctggtgttggctgattaatatgattagaccaaaaaatgctaaggactctacttctaatagt
atggagaacactgatagtacttggcctgaattgtttagagagttatgcaaaataaatgcatttgacactactgattcatcacttatgagaggcaaggagtttagtgactctatacataatacctttgactatatgtggagaaccaaggaacat
aatgaagttggttgattgctcctaagttctctggagaaagagatgaaagaaaatgatgatctcaggggatctgtctcccaccttcagaagcagatactgagccacaaatctgctaagattgccctgaatgagagttttaactcctgtagag

aaagagttgaaattgtgaaaaaacagagacaagctgttatcatgcgagtagctgatctgcaacaagaggtgcatgcacagccttgccaggtgtttactgttaaagtgagggcattgactggaaaaaaatgggaccctggaacttgga
gtggggatgtgtgggagaaccctgatgaagctgaggacactgagtttgtgaactctgatgaaactttttgccagaagaaacagtttccccatccccagtagtggtaacatcccctccctgacccgtgctgccattagcctttccaccttt
gtctgaggatgtaaaccctgcactgcttgaggcaacagtgatggccttccctgaggcagctgccaggcaagataatgttgattctcctcaagaggcacccctaatgcccctgaatgcttctagacctataactaggctaaattccttgcg
ggccccagaggtgaggttcagagtgtgacccatgaggaggtgcattatactctaaaagaactgcttaagctttctaatttatattggcagaaatctggagaacaggcatgggaatggatattaagggtaagggataatggtggaaggg
acatagagttggatcaagctgaatttattggtttggccctactaagtagggattctgcatttaatgttgcagctcggggacttagaaaaggttctgatagggccgggagcagtggctcacgcctgtaatcccagcaccttgggaggcgg
gggcgggcagatcacgagatcaggagattgagacaattctggctaaaatggtgaaaccccatctctgctaaaaatacaaaaattagctgggcatggtgatgcgtaactgtaatctcatctacttgggaggctgaggcaagagaactg
cttgaacctgtgaggcagagattgcagtgagccaagatcgccccactgcattccagcctggtaacagagcaagactccatttccaaaaaaaaaaaaaaaaaaagttataatagtttatttgcttggttagctgaaatatggattaaaagat
ggtccaatgttagtgagctggaaatgccttggtttaatgtagaggaagtgatccaaaggcttagggagattaggatggtggagtggattagtcactttagacctactcatcccagctgggagggtccagaagatacacccttggccga
agctttgtgaaatagatttgtgagagcagcacctgtatttttgaagagcccgtaattgctcttctctgtatgtcagatctaacagtaggaaccacagtcactcaactacaaaatttaaatacaatgggaataattggatcctgaggtggcag
gggccaagtgttggcactgaaccatcaaaggcaaggtgggcataactaccataatagacagcagaggcaaagcagccatcagaatagtctgactcatgtagagctctggcattggctaattaatcatggtgttcctagaagtgaaatt
gatgggaaacctactgtattcctacttgatttatataaacaaaaaactgccaggtagaatggactaaagactaatctgaattataaaaacagagaatcatgggccctcaatcaatttccagactcgaacctgttacagttccagaacccac
tgaatgaaggggaggctggatccccttgaggaaggacaccactaggctactgacaacttatgctgttactctttctcccatccttccctaaggagacctctggccttttaccagggtaactgtgtgtactggagaaagggaagtaatga
gacatttcagaaagtactggacactggctctgagctgacgttgattccagggtacccaaaacgttattgtggttccccagttaaagtaggggcttatggaggttaggtaattaatggagttttagctcatttctgacttacagtggttccagt
gggtccctggacttatcctctggtcattttcccagtgccaaaatgcataatttgtatagacatacttattagctggcagaaatgccacattggctccctgactggtaggatgagggctattatggtgggaaaggccaaacagaagccatta
gagctgtctctacctagaaaaataaaaaaatcaaaaacaatatcccatccctggagggactgaagtgattagtgtcaccatcaaggacttgaaagacgcaggggtggtgattcccaccacatccctgttcaactctcccatttgacctgt
gcagaggacagatggatcttggaaaatgatggtggattattttaagcttaaccaagtggtgactccaattgcagctgctctaccagttgtggttttgttgcttgagcaaattaacacatctcctggtgcctggtatgcagccattggcttggc
aagtggctttttctccattcctgtccataagacccaccagaagcaatttgccttcagctgacaaggccagcattatacctttaccaccctacctcaggggtgtatcaactctccagctttgtgtcataatcttatttggagagaccttgctcgc
ttttcacttccacgagatataacactggtccattacattcatgacattatgatgattggatacagtgagcaagaagtagcaaacacactgaacttattggtgagacatttgtatgccagaggatgggaaataaatccagctaaaatttaggg
actttctacctcggtaaaatttctagggttccagtggcatgagacctatggagatattccttctaaggtgaagcataacttgctgcgtttggcccctcttacaaccaagaaagaggcacaatgcctggtgggcctatttggattttggaggc
aacacattcctcgtttgggtgtgttactctggcccatttatcgagtgacctgaaaggctgccagatttaagtgcagtctagaacaaaagaaggctctgaaacaggtccaggctgctgtgaaagctgctctgccatttgggccacatgacc
ccgcagatccaatggtgcttgaggtgtcagtggcagatagggatgctgtttggagcctttggcaggcccccataggtgaatcacagtggagacctctaggattttggagcaaggccctgccacttctgcagataactactctcctttg
agagacagctattggtctgttattgggctttggtggtaactgaacgtttgactgtgggtcataaagtcaccatgctacctgaacctgcctatcatgaactggttgctttctgacccatctagccatgaagtgggtcagcacagcggcatttc
atcatcaaattgaagtggtgtgtatgtgatcgggcttgagcaggtcctgaaggcacaagtaagttacataaggaagtggctcaaatgcccatgttctccactcatgccaccctgccttccctcccccagcctgcaccaatggcctcatg
gggagttccctatgatcagttgacagaggaagggaagactaaggactggttcatagatggttctgcacgatatgcaggcaccacccgaaagtggacagctgcagcactatatccactttctaaatgcatgtgtacacttgtgctaaga
aaatatctttattttatttcctttatttttcctttatcatgtgaccttagatttatggacttcacatcagcatttaagcatttaagtgttgttcatatcagcatttaaatattgttaaccttatgtaataacttttggtttggggattggtgcgtttctggttgta
tgaggatagttgtattatattaggcataattatgaccttattattgtctttatttgaagattatgtatgatttcaggatgtgtgtatgggttcaagttgacaaggagttggacttgtgatggttaatactgtcaacttgattggattgaaagatgcaa
agtattaatctcggttatgtctgtgagggtgtggcaaaaggagattaacatttgagtcagtgggctgggaaggcagacccacccttaatctgggtacacaccatctaatcaagttccagtgtggccagattgtaaagcagggagaaaa
atgtgaaaagactagactgaattagcttcccagcctacatctttctcctgtgccaaatgcttcctgctcttgaacatcggactccaagttcttcagcgttgggagttggactggctttcttgctcctcagcttgcagagggcctgttgtggaa
ccttgtgatccgctgagttaatactacttaataagatcccctttatatacatataatatattatattatatataatatatataatatatattatatataatatatataatatattatatattatatataatatatattatatattatatataatatatattatatata
atatatattatatattatatattatatataatatatattatatataatatatataaaatatatatatatcctattagttctgtccctctagagaaccctgactaatacaatttatgtcattaatctcatttattgatttgtatacattgaaccaaccttatatccc
aggaataaaacctacttgattgtggtggattagctttttgatgtactcttggattcaattgctggtattttattgagaatttttgcatctgtgttcatcaaggatattggcttgaagttttcttttttgttgttccatatcagaatgatgacgacctcata
gaatgagttagtctgtcctcttttatcttttggaattgtttcaggaggcttgatatcagctcttctttatatgactggtatactttggctaggaatctctctggtccaggggtttttctggtgtaggtttttaattactgattcaacttcagaactcatta
ctcattattgagttctaaaactcactttcatgtactcttcaaaagactgtcttcttctgttgttgagcggggtgttctctcaaggtcgtttaggtgaaggtggttgctggtgttcttctgtatccttactgcttgtctttctcttttttattgactactga
ggattaatggtgatgtgtccaactttaactctagattagtctatttctcttttagattgtaactctgttttatatattttgaagctctgttgttaggcatgtgtatttggattgttaggtcttcttgatgatgacctttatcattatgtaatgtttcttcttatct
ctggaagtattcgttgttctgaagtctatttgtgctgatatgaatacagccttcacagctctattttcactagtatttgtatatctttttctcagcttttaaattgagatgttcagaccatttgcattaaagtagttgttaataggattaaatttaaatcta

ccattaagtggttatttctcttgtccatttaaacttgttcctttttcatatttttctgccttcatttatattgagtttatctccacgacttacttattaaattaatttttaatggtttagtatttccacaatgtttatatatatacttgatttttcacattc
caccttcaaatgacagaattatactggatatatagaaatcttacatcattgcacttcctcctcctcaaaatgttgtgctatttgtctcttgtaatagaggcttacttctattatgttatagctctcatataatacattgacactatttttaccctgaa
taatcagttgttttttaaagtgattatgactacaaatatttgaataatttctttatttaccatttcctgtgctctcttacagtagatcccaatttccatctgtgaaaaacaacctttagcatttcttatagcacgggga
ctgctgttgctgttgtctttcagctttctttgtctgaagaagtctttatttgcctcagttttaaaagtgatttgctgagtatagatactgggttgagagttttcatttcctgtatcatttaacaatgatgttccattatattccgtttgaatagtttct
gactagaaatctgatctttgtttcttgtattcaatagttcctttttctctgactgccttaagatatttctcatctttgtcttgactatataatttgtattattaacttttgtatttattctgcttgagtttcctgagctcctggatttgcagatt
gttgatttttattgttttgtaaaattcatagccattatctattctactgttttgttttttttcacttctctctctctctgtattcttctttttggactgtaagtattcaaatgttagatcattcaatattgcttcttctgctttttttttttgtc
aggaactctttttgtatctgtgttggttggataagttctagtagactatgttcaagtttattttgttagttgtgtctaatttgactcctcagtgcattcagagaatttcttcatctctgatattataaatctcttcctagcatttttcatgttactct
ttctatagtttccatctcttttgctgaaattctcccccctatccatggatattgtccacctttaccacaaagattctttaacatattaacataggtatcatacaaacccaaactgatagtttccagatggtgtctttctgagtctgtctgcttgattgctt
tattatttaacagtgacttatcttcccttcttcagctttttggtgtgtctgttaattgtttaatcaaacactgggtatcataaaatggaggaaacagtagagattgcagtaaatattatttatgctttgaaatgggcaccccatcttctgttgaaaatatgtttt
gtggtcaattgagtcaacctagtaactggttgaactgaatttggcatttgtgctgttgtttaaatgcaccacaggtttaaattcctccagtgatgggttgctgctatctttgcttagatgtgggctgtgggtgtggaagaatttctc
agtgttcctatctattattagatttagcagtcactgcatgcctgcactacagaggggatatcttcatacacataatctaacccccattgaaactgctgttcttcttaatgaatgctcaatcttttggtggaaataaacaaatgctgtatctcctgga
gccacttcagtcttagtcaggttctgcaggggcttgaaggaatgcatttctcagtattcttgtgcctttattggatggaactttgaacctgtgtgggtttggagagaaagagtagcagacgtctgctatgttgcaatgcaggatgctgggc
acaagaaaatttccagtctctcctccaaggaaataagatttgatcatctacctatcctgagaagtgaagggctttgcctgcggtgctagatgcaaaaccatttttctcccccccattgcccagaaacttaaggctttggctttctgagcagt
ggtctaggaattgtgcaaggtttcatatttgacccctgacagcccatcaccacctacagcttgcagtgccaaatgtatctccctctgatctctcctgtctgtggtcctcatgaacattaagaagagatttctaaaaaagagctttgcacatg
agcatagtttctggtgagaagaattctgatatgttaacttcctctaaacttttaaataaaatatttctaagaattaaatgaatctatttcctttggttttttgcacgtctgtctgcctaatcaagagaagaagaatggt
cgtaattctcagagacttttttcctgtttgtgtcataaatgacttcacatttttttctgttctaagaactattcagcttgatttcttctgttttaatttttagCAGCACCTGAGCAAAGCCATGTGGTCCAGGATTGCTAC
CATGGTGATGGACAGAGTTATCGAGGCACGTACTCCACCACTGTCACAGGAAGGACCTGCCAAGCTTGGTCATCTATGACACCCATCAACATAATA
GGACCACAGAAAACTACCCAAATGC (SEQ ID NO: 1805)

Exon 1-2 of *Pcsk9*

AGCGACGTCGAGGCGCTCATGGTTGCAGGCGGGCGCCGCCGTTCAGTTCAGGGTCTGAGCCTGGAGGAGTGAGCCAGGCAGTGA
GACTGGCTCGGGCGGGCCGGGACGCGTCGTTGCAGCAGCGGCTCCCAGCTCCCAGGCCAGGATTCCGGCGGCGCCCCTTCACGCGCCC
TGCTCCTGAACTTCAGCTCCTGCACAGTCCTCCCCACCGCAAGGCTCAAGGCGCCCGCCGGCGTGGACCGCGCACGGCCTCTAGGT
CTCCTCGCCAGGACAGCAACCTCTCCCCTGGCCCTCATGGGCACCGTCAGCTCCAGGCGGTCCTGGTGGCCGCTGCCACTGCTGC
TGCTGCTGCTGCTCCTGGGTCCCGCGGCGCCCGTGCGCAGGAGGACGGCGACTACGAGGAGCTGGTGCTAGCCT
TGCGTTCCGAGGAGGACGGCCTGGCCGAAGCACCCGAGCACGGAACCACAGCCACCTTCCACCGCTGCGCCAAGgtgcgggtgtagggatgg
gaggccgggggcgaacccgcagccgggacgggtgcggtgctgtttcctctcgggcctcagtttcccccatgtaagagaggaagtggagtgcaggtcgccgagggctcttcgcttggcacgatcttgggggactgcaggcaaggccg
gcgggggaggacgggtagtgggggagcacggtggagagcgggacggccggctctttgggggacttgctgggggcgtgcggctgcgctattcagtgggaaggttcgcgggggttgggagacccggaggccgaggaaggcga
gcagagcactgccaggatatcctgcccagatttccctgcctgccgcggcacaggtggtgaaggagtgaatgcctggaaacgtactgggaaactgcaccaggcacagagaaaagcgggcttgccattatagtgggttccga
ttggtttggaaaacatgggcagcggaggtggagggcctggagagaagggccctaccgagacagggcgggtgggaaggacgggcagatgctgggagcacgaggcaatttctttatgacacagaactcatgctctagtattc
catctgtttcagccgaagaaaagaaccagctgaagggcagggggagaagggcggaggtattctcgaggcccattggcgtccttaggactcaggcagggaaggcccttggtgctctggagccggagtggtgcgccctggta
ctgggacccggagctgagcccggcgcctcagcccacctggctgtctgccgaccgtgtgcgggcgagtttgctcaacaactctgccagcttctggcctcaggctgtggaagcttcttccgggggcgagaccactagcttttc
taagtattaccagcccaggacttggctgaggttctgtgtcccccagcttggagtcaagatgtgggttgaatcttggcttcctctcactagctgtggtgcttgacaagtcacttatccttgagcctccattgcctaatctttaaaagggaggtg
acaatcgtccctacggctcagtggcagcagatgggggagatgaaggaaagttctgttgaccatgagtgaacttacaatgcaagcccgggggggatcactgcagtttgtccctgtctgcagtgtgacctgttggtgacattgtcttg
ctccaaaccacagctcctgggcagagggaaaattctgccactcacagctgcctgcccacgcttctgtctgagtgtgctgggtggcaggatggcaagtccttactcagctcagtatagccctcttccttgttccctgagccttgacct

tctcgagggatgttgtgggttgtggccaggataagaaagggcatttcaagttaccactgctccaaaacaactgttctggaaatagtgagtaccccatcctgagaggtgagtaagcagaggctgtatgaccacctgaaccaagccctt
gaggatgtttcttctctgttgtggaagtttggaacaggaggagcctcctcaagttcatttatttcattcaattcaattttagaatgaaaatatttttggcaagcagaaaaatatttttagaccaatcctttctttttagtcatg
agaaactgaggcccagagagaggaggtcaccccagtgcattagaactgggtttccagaactcgacactcccactgcacagagtactctcccaattcattcaattttattagcggaaggcatttcagatggtcttgaagcattagta
ggagttcagcgatgatggtgtcatgagaatttattctaggattaggaggtaccatgaacaaagatacagagctgggaaaaaccagaggtggaagataaggagcacatgtccacagttcttttctttttttttgagatggagtttcgctct
gttgcccaggctggagtgcaatggtgcagtctcagctcactgcaacatctgtcctcccgggttcaagtgattctcctgcctcagcctcccaagaagctgggattacaggtaccctgccaccacgcccggctaatttttgtatttttagtagag
aaggggttcaccacgttggccaggctagtcgcaaactcctgacctcctcagtggatccgccgcctcagcctcccaaagtgctcgaattacaggtgtgagccaccacgcctggcctccacagttctttatccac
cgtctgaaatgtaaaatgttacgaaaaccaaaagttttttttgtgatttatttgatgtagcacctgacgtgaactgacatgagattatttttaatttagttgtgtgaatatgcattctatatatttgctgcatagattacagtatgcagctccaga
ttcttccaagcagactctgattgcccattactgccttctaaaatccaaacaagttctgaggttcaaaaccgttttggcctaaggcttgggtaaaggggtggactcttctactctgactggagtccaagatgcatatatacagagata
tgggtgatgggctgcaaggtaggttgaggtagggccaaggagaggagcatggagttttgaattcatgaggctgtgggagccagtgaaggttcttaagcaggtatgtctgcctgagagcagacaagaagctaa
aaaccaaacaaatcaccatagatatggctgctataatttgtttgtcccctccaaatctcatgtggaaatttggtcctcagtgttgaagtgtttgggtcatgggggaggaaccccctgtgaaaggcttggtgc
cgtccttgtgataatgagtaagttctcccgctatgattccctgaaggctgattattaaaaagagctggcacctccctcttctctctgcttcttctctgccatgtagattgatctctgcacatgtaggctcccctcacttcatgccatcagt
gaaagcagcttaaggccctcaccagaaagcagatgctggtgccatgcttcctggagaagcttgcagaatcatgagctgaataaatcccttttccttgtaaattactcaccttcaggtattccttatatagcaacacacaaaaggactaagacag
tggcctgacttttctctctctttaagaagtgttgccttgctcacttagtcatcccttctgcctgcatttgtagagcatctgagatgggagatttatatacccgtcactcttgacttcccagcaggcctatgtcatagtgtactgtggtctctacaat
acagcagaggtatctgaggctccgagaggttgagtgacttgctcatggctgcacaaccagtaaatattggagctggaattcaggtccacggtttcctggctccaaagcccatgatttttcctcaatttattctgactgggcatgggg
gaggggtggccttgggcagggccaccaggagcgaccagggcccgtagagagctgggtgcaggtacagaggaaaacctgttgtcgagtgtggcccgtagttcccattttgcctgaatggcacatttgaaagtgttatataaccat
gtgaataataatagttggcctatatgagttctttaaattgctttttggtccgcatttgtgtaacttctttatcatctactactatctgttgtgtctctttgttaattgtaagtagggtgagatataaagtacacctagggtttgctgggttctccatg
tcatcatgttcctccttgcatgggggccag*GATCCGTGGAGGTTGCCTGGCACCTACGTGGTGTGGCTGAAGGAGGAGACCCACCTCTCGCAGTCAGAGCGCACTGC*
CCGCCGCCTGCAGGCCCAGGCTGCCCCGCCGGGGATACCTCCTGCATGTCTTCCATGGCCTTCTTCCTGGCTTCCTGGTGAAGATGAG
TGGCGACCTGCTGGAGCTG (**SEQ ID NO: 1806**)

Exon 1-2 of TF

ACAGAAGCGGAGTCCGACTGTGCTCGCTGCTCAGCGCCGCCGCACCCGGGAAGATGAGGCTCCTGCTGGTCTGGTCTGCGCCG
TCCTGGgtgagtgcgggcacgggggtagcaccgcagagtcgctggcccgcgcgttccctgcaacccgggcggcaccgcgcagcctgcatgcactccgcgctcaggctggaagcctgggtgtctgggtgccttccattg
cctctctacgccccacccctgggcttgcttgggcttctagacaccttcactgcactcactgtcttctccccctcctcctcttgtccagtaaatcttgcagcagtgaaatacaattaaatcagagcacgtctaaccaggttatcaaaaccctt
tgctgctgcctcttgtcaagtccgaactcctagcatggcattccaaggcctcccgggtctgaccattcccctcctgggtgcccctgggtgccatcgagcggtca
gagcatgggtttgaagccagacttcttggatccaagtcctggccggctcctcaccagggagggggtcatcacagcacttgcctgggaggtgcagcgaggtggcagatgctgagtaccagtggcagatgttgc
acacatcctgctatgggcagtggctgacgccacactcccgacctcctgaggtgggccccccttcatactctaacttctgcctgccattcatgggggcatttgtcacactgtttgggcttccacgtttagtttctctcccataagctagc
aattccttgagaggaggacccccacgttactgccttatccccgcacagaggctgtgcaaggtaatgctccactgaggacacattctgccctatgggaactctggaggcagggcttgtggaggggctgccacaggctt
atgttgcctctaggatttccatgggccaagagggaaaatgggggtcgctgggtggccatcctggtgccccctcctcatgcatcctggagagctgtgggcctcctctccacaggcttttgtggatcttcttcagagccagtgagtc
agccttgctgtggtggcccacaaggagtgagatgacagagaggccttgggacaggcaaaagctcatgtgataggatgctgtctctgcaacccacatggatttaatagttacccatggcttcagcttttttgtgaatggcag
ccaaaaagtggtggtgaaaatggagggatagttcagtggagacagaggggagactgttgaatttgtggagttttagttcttcctaaacagagagcaacagacaagagttacagacaaattaatttcttttcgttaatttaccctcaactactg
gccaagaagggggtgtaactcagaatgtttttcttctttgtgccctgagtctgagtagcagcacccagattctaggcctggcttgggccaacgacacaagcagggtgacctttgggttaaaaaaaaaaaaaaggcatcattctctg
agtctcactcctcttctcaaaaagatggcaattcctccccccgctggtttccatagttgaggtagcaagccaatgttgggggtagctgaaaacgccctgtgcatactgaggcttatgttccatgggggggccaggggccaggacagtg
atgggaggagacaaggcggatacagaggagcagattgtcatctccagctgggaggtgatggccatgcctgcaccctctggccagcagagggtggtcagtaggaaactgggaggccattagggcaacctctattggctcagac
tcagaatgcagtagaacttgtgccctgtagtgttcatggacaggagtgagaggaggacaggactgaggggatgtggctgtcaaggccttctagggggcgatgctgtctctccctcagcataggggagtgggccctccacctctggcc

tctctcccccag*GGCTGTGTCTGGCTGTCCCTGATAAAACTGTGAGATGGTGTGCAGTGTCGGAGCATGAGGCCACTAAGTGCCAGAGTTTCCGCGACCAT ATGAAAAGCGTCATTCCATCCGATGGTCCCAGTGTTGCTTGTGTGAAGAAAGCCTCCTACCTTGATTGCATCAGGGCCATTGCG* (**SEQ ID NO: 1807**)
Exon 1-2 of TTR
ACAGAAGTCCACTCATTCTTGGCAGGATGGCTTCTCATCGTCTGCTCCTCCTCTGCCTTGCTGGACTGGTATTTGTGTCTGAGGCT GGCCCTACGgtgagtgtttctgtgacatcccattcctacatttaagattcacgctaaatgaagtagaagtgactccttccagctttgccaaccagcttttattactagggcaagggtacccagcatctatttttaatataattaattca aacttcaaaaagaatgaagttccactgagcttactgagctgggacttgaactctgagcattctacctcattgctttggtgcattaggtttgtaatatctggtacctctgtttcctcagatagatgatagaaataaagatatgatattaaggaag ctgttaatactgaattttcagaaaagtatccctccataaaatgtatttgggggacaaactgcaggagattatattctggccctatagttattcaaaacgtatttattgattaatctttaaaaggcttagtgaacaatattctagtcagatatctaatt cttaaatcctctagaagaattaactaatactataaaatgggtctggatgtagttctgacattattttataacaactggtaagagggagtgactatagcaacaactaaaatgatctcaggaaaacctgtttggccctatgtatggtacattacat cttttcagtaattccactcaaatggagacttttaacaaagcaactgttctcaggggacctattttctcccttaaaattcattatacacatccctggttgatagcagtgtgtctggaggcagaaaccattcttgctttggaaacaattacgtctgtg ttatactgagtagggaagctcattaattgtcgacacttacgttcctgataatgggatcagtgtgtaattcttgtttcgctccagatttctaataccacaaagaataaatcctttcactctgatcaattttgttaacttctcacgtgtcttctctacacc cag*GGCACCGGTGAATCCAAGTGTCCTCTGATGGTCAAAGTTCTAGATGCTGTCCGAGGCAGTCCTGCCATCAATGTGGCCGTGCATGTGTTCAGAAA GGCTGCTGATGACACCTGGGAGCCATTTGCCTCTGG* (**SEQ ID NO: 1808**)

**Uppercase text represents entire gene; uppercase, bold text represents exon 1 sequence; lowercase text represents intron 1 sequence; uppercase, italic text represents exon 2 sequence.*

[687] **TABLE 11** provides exemplary functional human protein sequences.

TABLE 11. Exemplary Functional Human Proteins

SEQ ID NO: (Protein)	Sequence
269 (GAA)	MGVRHPPCSHRLLAVCALVSLATAALLGHILLHDFLLVPRELSGSSPVLEETHPAHQQGASRPGPRDAQAHPGRPRAVPTQCDV PPNSRFDCAPDKAITQEQCEARGCCYIPAKQGLQGAQMGQPWCFFPPSYPSYKLENLSSSEMGYTATLTRTTPTFFPKDILTLRL DVMMETENRLHFTIKDPANRRYEVPLETPHVHSRAPSPLYSVEFSEEPFGVIVRRQLDGRVLLNTTVAPLFFADQFLQLSTSLPS QYITGLAEHLSPLMLSTSWTRITLWNRDLAPTPGANLYGSHPFYLALEDGGSAHGVFLLNSNAMDVVLQPSPALSWRSTGGILD VYIFLGPEPKSVVQQYLDVVGYPFMPPYWGLGFHLCRWGYSSTAITRQVVENMTRAHFPLDVQWNDLDYMDSRRDFTFNKD GFRDFPAMVQELHQGGRRYMMIVDPAISSSGPAGSYRPYDEGLRRGVFITNETGQPLIGKVWPGSTAFPDFTNPTALAWWEDM VAEFHDQVPFDGMWIDMNEPSNFIRGSEDGCPNNELENPPYVPGVVGGTLQAATICASSHQFLSTHYNLHNLYGLTEAIASHRA LVKARGTRPFVISRSTFAGHGRYAGHWTGDVWSSWEQLASSVPEILQFNLLGVPLVGADVCGFLGNTSEELCVRWTQLGAFYP FMRNHNSLLSLPQEPYSFSEPAQQAMRKALTLRYALLPHLYTLFHQAHVAGETVARPLFLEFPKDSSTWTVDHQLLWGEALLIT PVLQAGKAEVTGYFPLGTWYDLQTVPVEALGSLPPPPAAPREPAIHSEGQWVTLPAPLDTINVHLRAGYIIPLQGPGLTTTESRQ QPMALAVALTKGGEARGELFWDDGESLEVLERGAYTQVIFLARNNTIVNELVRVTSEGAGLQLQKVTVLGVATAPQQVLSNG VPVSNFTYSPDTKVLDICVSLLMGEQFLVSWC

SEQ ID NO: (Protein)	Sequence
270 (CFTR)	MQRSPLEKASVVSKLFFSWTRPILRKGYRQRLELSDIYQIPSVDSADNLSEKLEREWDRELASKKNPKLINALRRCFFWRFMFY GIFLYLGEVTKAVQPLLLGRIIASYDPDNKEERSIAIYLGIGLCLLFIVRTLLLHPAIFGLHHIGMQMRIAMFSLIYKKTLKLSSRVL DKISIGQLVSLLSNNLNKFDEGLALAHFVWIAPLQVALLMGLIWELLQASAFCGLGFLIVLALFQAGLGRMMMKYRDQRAGKI SERLVITSEMIENIQSVKAYCWEEAMEKMIENLRQTELKLTRKAAYVRYFNSSAFFFSGFFVVFLSVLPYALIKGIILRKIFTTISFC IVLRMAVTRQFPWAVQTWYDSLGAINKIQDFLQKQEYKTLEYNLTTTEVVMENVTAFWEEGFGELFEKAKQNNNNRKTSNGD DSLFFSNFSLLGTPVLKDINFKIERGQLLAVAGSTGAGKTSLLMVIMGELEPSEGKIKHSGRISFCSQFSWIMPGTIKENIIFGVSYD EYRYRSVIKACQLEEDISKFAEKDNIVLGEGGITLSGGQRARISLARAVYKDADLYLLDSPFGYLDVLTEKEIFESCVCKLMANK TRILVTSKMEHLKKADKILILHEGSSYFYGTFSELQNLQPDFSSKLMGCDSFDQFSAERRNSILTETLHRFSLEGDAPVSWTETKK QSFKQTGEFGEKRKNSILNPINSIRKFSIVQKTPLQMNGIEEDSDEPLERRLSLVPDSEQGEAILPRISVISTGPTLQARRRQSVLNL MTHSVNQGQNIHRKTTASTRKVSLAPQANLTELDIYSRRLSQETGLEISEEINEEDLKECFFDDMESIPAVTTWNTYLRYITVHKS LIFVLIWCLVIFLAEVAASLVVLWLLGNTPLQDKGNSTHSRNNSYAVIITSTSSYYVFYIYVGVADTLLAMGFFRGLPLVHTLITV SKILHHKMLHSVLQAPMSTLNTLKAGGILNRFSKDIAILDDLLPLTIFDFIQLLLIVIGAIAVVAVLQPYIFVATVPVIVAFIMLRAY FLQTSQQLKQLESEGRSPIFTHLVTSLKGLWTLRAFGRQPYFETLFHKALNLHTANWFLYLSTLRWFQMRIEMIFVIFFIAVTFISI LTTGEGEGRVGIILTLAMNIMSTLQWAVNSSIDVDSLMRSVSRVFKFIDMPTEGKPTKSTKPYKNGQLSKVMIIENSHVKKDDI WPSGGQMTVKDLTAKYTEGGNAILENISFSISPGQRVGLLGRTGSGKSTLLSAFLRLLNTEGEIQIDGVSWDSITLQQWRKAFGV IPQKVFIFSGTFRKNLDPYEQWSDQEIWKVADEVGLRSVIEQFPGKLDFVLVDGGCVLSHGHKQLMCLARSVLSKAKILLLDEP SAHLDPVTYQIIRRTLKQAFADCTVILCEHRIEAMLECQQFLVIEENKVRQYDSIQKLLNERSLFRQAISPSDRVKLFPHRNSSKC KSKPQIAALKEETEEEVQDTRL
271 (DMD)	MLWWEEVEDCYEREDVQKKTFTKWVNAQFSKFGKQHIENLFSDLQDGRRLLDLLEGLTGQKLPKEKGSTRVHALNNVNKAL RVLQNNNVDLVNIGSTDIVDGNHKLTLGLIWNIILHWQVKNVMKNIMAGLQQTNSEKILLSWVRQSTRNYPQVNVINFTTSWS DGLALNALIHSHRPDLFDWNSVVCQQSATQRLEHAFNIARYQLGIEKLLDPEDVDTTYPDKKSILMYITSLFQVLPQQVSIEAIQ EVEMLPRPPKVTKEEHFQLHHQMHYSQQITVSLAQGYERTSSPKPRFKSYAYTQAAYVTTSDPTRSPFPSQHLEAPEDKSFGSSL MESEVNLDRYQTALEEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTAHQGRVGNILQLGSKLIGTGKLSEDE ETEVQEQMNLLNSRWECLRVASMEKQSNLHRVLMDLQNQKLKELNDWLTKTEERTRKMEEEPLGPDLEDLKRQVQQHKVLQ EDLEQEQVRVNSLTHMVVVVDESSGDHATAALEEQLKVLGDRWANICRWTEDRWVLLQDILLKWQRLTEEQCLFSAWLSEK EDAVNKIHTTGFKDQNEMLSSLQKLAVLKADLEKKKQSMGKLYSLKQDLLSTLKNKSVTQKTEAWLDNFARCWDNLVQKLE KSTAQISQAVTTTQPSLTQTTVMETVTTVTTREQILVKHAQEELPPPPQKKRQITVDSEIRKRLDVDITELHSWITRSEAVLQSPE FAIFRKEGNFSDLKEKVNAIEREKAEKFRKLQDASRSAQALVEQMVNEGVNADSIKQASEQLNSRWIEFCQLLSERLNWLEYQ NNIIAFYNQLQQLEQMTTTAENWLKIQPTTPSEPTAIKSQLKICKDEVNRLSDLQPQIERLKIQSIALKEKGQGPMFLDADFVAFT NHFKQVFSDVQAREKELQTIFDTLPPMRYQETMSAIRTWVQQSETKLSIPQLSVTDYEIMEQRLGELQALQSSLQEQQSGLYYL STTVKEMSKKAPSEISRKYQSEFEEIEGRWKKLSSQLVEHCQKLEEQMNKLRKIQNHIQTLKKWMAEVDVFLKEEWPALGDSEI LKKQLKQCRLLVSDIQTIQPSLNSVNEGGQKIKNEAEPEFASRLETELKELNTQWDHMCQQVYARKEALKGGLEKTVSLQKDL

SEQ ID NO: (Protein)	Sequence
	SEMHEWMTQAEEEYLERDFEYKTPDELQKAVEEMKRAKEEAQQKEAKVKLLTESVNSVIAQAPPVAQEALKKELETLTTNYQ
	WLCTRLNGKCKTLEEVWACWHELLSYLEKANKWLNEVEFKLKTTENIPGGAEEISEVLDSLENLMRHSEDNPNQIRILAQTLT
	DGGVMDELINEELETFNSRWRELHEEAVRRQKLLEQSIQSAQETEKSLHLIQESLTFIDKQLAAYIADKVDAAQMPQEAQKIQS
	DLTSHEISLEEMKKHNQGKEAAQRVLSQIDVAQKKLQDVSMKFRLFQKPANFEQRLQESKMILDEVKMHLPALETKSVEQEVV
	QSQLNHCVNLYKSLSEVKSEVEMVIKTGRQIVQKKQTENPKELDERVTALKLHYNELGAKVTERKQQLEKCLKLSRKMRKEM
	NVLTEWLAATDMELTKRSAVEGMPSNLDSEVAWGKATQKEIEKQKVHLKSITEVGEALKTVLGKKETLVEDKLSLLNSNWIA
	VTSRAEEWLNLLLEYQKHMETFDQNVDHITKWIIQADTLLDESEKKKPQQKEDVLKRLKAELNDIRPKVDSTRDQAANLMAN
	RGDHCRKLVEPQISELNHRFAAISHRIKTGKASIPLKELEQFNSDIQKLLEPLEAEIQQGVNLKEEDFNKDMNEDNEGTVKELLQ
	RGDNLQQRITDERKREEIKIKQQLLQTKHNALKDLRSQRRKKALEISHQWYQYKRQADDLLKCLDDIEKKLASLPEPRDERKIK
	EIDRELQKKKEELNAVRRQAEGLSEDGAAMAVEPTQIQLSKRWREIESKFAQFRRLNFAQIHTVREETMMVMTEDMPLEISYVP
	STYLTEITHVSQALLEVEQLLNAPDLCAKDFEDLFKQEESLKNIKDSLQQSSGRIDIIHSKKTAALQSATPVERVKLQEALSQLDF
	QWEKVNKMYKDRQGRFDRSVEKWRRFHYDIKIFNQWLTEAEQFLRKTQIPENWEHAKYKWYLKELQDGIGQRQTVVRTLNA
	TGEEIIQQSSKTDASILQEKLGSLNLRWQEVCKQLSDRKKRLEEQKNILSEFQRDLNEFVLWLEEADNIASIPLEPGKEQQLKEKL
	EQVKLLVEELPLRQGILKQLNETGGPVLVSAPISPEEQDKLENKLKQTNLQWIKVSRALPEKQGEIEAQIKDLGQLEKKLEDLEE
	QLNHLLLWLSPIRNQLEIYNQPNQEGPFDVKETEIAVQAKQPDVEEILSKGQHLYKEKPATQPVKRKRLEDLSSEWKAVNRLLQE
	LRAKQPDLAPGLTTIGASPTQTVTLVTQPVVTKETAISKLEMPSSLMLEVPALADFNRAWTELTDWLSLLDQVIKSQRVMVGDL
	EDINEMIIKQKATMQDLEQRRPQLEELITAAQNLKNKTSNQEARTIITDRIERIQNQWDEVQEHLQNRRQQLNEMLKDSTQWLE
	AKEEAEQVLGQARAKLESWKEGPYTVDAIQKKITETKQLAKDLRQWQTNVDVANDLALKLLRDYSADDTRKVHMITENINAS
	WRSIHKRVSEREAALEETHRLLQQFPLDLEKFLAWLTEAETTANVLQDATRKERLLEDSKGVKELMKQWQDLQGEIEAHTDV
	YHNLDENSQKILRSLEGSDDAVLLQRRLDNMNFKWSELRKKSLNIRSHLEASSDQWKRLHLSLQELLVWLQLKDDELSRQAPI
	GGDFPAVQKQNDVHRAFKRELKTKEPVIMSTLETVRIFLTEQPLEGLEKLYQEPRELPPEERAQNVTRLLRKQAEEVNTEWEKL
	NLHSADWQRKIDETLERLQEATDELDLKLRQAEVIKGSWQPVGDLLIDSLQDHLEKVKALRGEIAPLKENVSHVNDLARQ
	LTTLGIQLSPYNLSTLEDLNTRWKLLQVAVEDRVRQLHEAHRDFGPASQHFLSTSVQGPWERAISPNKVPYYINHETQTTCWDH
	PKMTELYQSLADLNNVRFSAYRTAMKLRRLQKALCLDLLSLSAACDALDQHNLKQNDQPMDILQIINCLTTIYDRLEQEHNNL
	VNVPLCVDMCLNWLLNVYDTGRTGRIRVLSFKTGIISLCKAHLEDKYRYLFKQVASSTGFCDQRRLGLLLHDSIQIPRQLGEVA
	SFGGSNIEPSVRSCFQFANNKPEIEAALFLDWMRLEPQSMVWLPVLHRVAAAETAKHQAKCNICKECPIIGFRYRSLKHFNYDIC
	QSCFFSGRVAKGHKMHYPMVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYLPVQTVLEGDNMETPVTLINFWPV
	DSAPASSPQLSHDDTHSRIEHYASRLAEMENSNGSYLNDSISPNESIDDEHLLIQHYCQSLNQDSPLSQPRSPAQILISLESEERGEL
	ERILADLEEENRNLQAEYDRLKQQHEHKGLSPLPSPPEMMPTSPQSPRDAELIAEAKLLRQHKGRLEARMQILEDHNKQLESQL
	HRLRQLLEQPQAEAKVNGTTVSSPSTSLQRSDSSQPMLLRVVGSQTSDSMGEEDLLSPPQDTSTGLEEVMEQLNNSFPSSRGRN
	TPGKPMREDTM

SEQ ID NO: (Protein)	Sequence
272 (FXN)	MWTLGRRAVAGLLASPSPAQAQTLTRVPRPAELAPLCGRRGLRTDIDATCTPRRASSNQRGLNQIWNVKKQSVYLMNLRKSG TLGHPGSLDETTYERLAEETLDSLAEFFEDLADKPYTFEDYDVSFGSGVLTVKLGGDLGTYVINKQTPNKQIWLSSPSSGPKRYD WTGKNWVYSHDGVSLHELLAAELTKALKTKLDLSSLAYSGKDA
273 (F8)	MQIELSTCFFLCLLRFCFSATRRYYLGAVELSWDYMQSDLGELPVDARFPPRVPKSFPFNTSVVYKKTLFVEFTDHLFNIAKPRP PWMGLLGPTIQAEVYDTVVITLKNMASHPVSLHAVGVSYWKASEGAEYDDQTSQREKEDDKVFPGGSHTYVWQVLKENGPM ASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVFDEGKSWHSETKNSLMQDRDAASARAWPK MHTVNGYVNRSLPGLIGCHRKSVYWHVIGMGTTPEVHSIFLEGHTFLVRNHRQASLEISPITFLTAQTLLMDLGQFLLFCHISSH QHDGMEAYVKVDSCPEEPQLRMKNNEEAEDYDDDLTDSEMDVVRFDDDNSPSFIQIRSVAKKHPKTWVHYIAAEEEDWDYA PLVLAPDDRSYKSQYLNNGPQRIGRKYKKVRFMAYTDETFKTREAIQHESGILGPLLYGEVGDTLLIIFKNQASRPYNIYPHGIT DVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRYYSSFVNMERDLASGLIGPLLICYKESVDQRGNQI MSDKRNVILFSVFDENRSWYLTENIQRFLPNPAGVQLEDPEFQASNIMHSINGYVFDSLQLSVCLHEVAYWYILSIGAQTDFLSV FFSGYTFKHKMVYEDTLTLFPFSGETVFMSMENPGLWILGCHNSDFRNRGMTALLKVSSCDKNTGDYYEDSYEDISAYLLSKN NAIEPRSFSQNSRHPSTRQKQFNATTIPENDIEKTDPWFAHRTPMPKIQNVSSSDLLMLLRQSPTPHGLSLSDLQEAKYETFSDDP SPGAIDSNNSLSEMTHFRPQLHHSGDMVFTPESGLQLRLNEKLGTTAATELKKLDFKVSSTSNNLISTIPSDNLAAGTDNTSSLGP PSMPVHYDSQLDTTLFGKKSSPLTESGGPLSLSEENNDSKLLESGLMNSQESSWGKNVSSTESGRLFKGKRAHGPALLTKDNAL FKVSISLLKTNKTSNNSATNRKTHIDGPSLLIENSPSVWQNILESDTEFKKVTPLIHDRMLMDKNATALRLNHMSNKTTSSKNME MVQQKKEGPIPPDAQNPDMSFFKMLFLPESARWIQRTHGKNSLNSGQGPSPKQLVSLGPEKSVEGQNFLSEKNKVVVGKGEFT KDVGLKEMVFPSSRNLFLTNLDNLHENNTHNQEKKIQEEIEKKETLIQENVVLPQIHTVTGTKNFMKNLFLLSTRQNVEGSYDG AYAPVLQDFRSLNDSTNRTKKHTAHFSKKGEEENLEGLGNQTKQIVEKYACTTRISPNTSQQNFVTQRSKRALKQFRLPLEETE LEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKEKGAITQSPLSDCLTRSHSIPQANRSPLPIAKVSSFPSIRPIYLTRVLFQDNSS HLPAASYRKKDSGVQESSHFLQGAKKNNLSLAILTLEMTGDQREVGSLGTSATNSVTYKKVENTVLPKPDLPKTSGKVELLPK VHIYQKDLFPTETSNGSPGHLDLVEGSLLQGTEGAIKWNEANRPGKVPFLRVATESSAKTPSKLLDPLAWDNHYGTQIPKEEWK SQEKSPEKTAFKKKDTILSLNACESNHAIAAINEGQNKPEIEVTWAKQGRTERLCSQNPPVLKRHQREITRTTLQSDQEEIDYDD TISVEMKKEDFDIYDEDENQSPRSFQKKTRHYFIAAVERLWDYGMSSSPHVLRNRAQSGSVPQFKKVVFQEFTDGSFTQPLYRG ELNEHLGLLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLISYEEDQRQGAEPRKNFVKPNETKTYFWKVQHHMAPTKDEFDCK AWAYFSDVDLEKDVHSGLIGPLLVCHTNTLNPAHGRQVTVQEFALFFTIFDETKSWYFTENMERNCRAPCNIQMEDPTFKENY RFHAINGYIMDTLPGLVMAQDQRIRWYLLSMGSNENIHSIHFSGHVFTVRKKEEYKMALYNLYPGVFETVEMLPSKAGIWRVE CLIGEHLHAGMSTLFLVYSNKCQTPLGMASGHIRDFQITASGQYGQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIIHGI KTQGARQKFSSLYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKHNIFNPPIIARYIRLHPTHYSIRSTLRMELMGCD LNSCSMPLGMESKAISDAQITASSYFTNMFATWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFQKTMKVTGVTTQGVKSLL TSMYVKEFLISSSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWVHQIALRMEVLGCEAQDLY

SEQ ID NO: (Protein)	Sequence
274 (F9)	MQRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNRPKRYNSGKLEEFVQGNLERECMEEKCSFEEAREVFENTERT TEFWKQYVDGDQCESNPCLNGGSCKDDINSYECWCPFGFEGKNCELDVTCNIKNGRCEQFCKNSADNKVVCSCTEGYRLAEN QKSCEPAVPFPCGRVSVSQTSKLTRAETVFPDVDYVNSTEAETILDNITQSTQSFNDFTRVVGGEDAKPGQFPWQVVLNGKVDA FCGGSIVNEKWIVTAAHCVETGVKITVVAGEHNIEETEHTEQKRNVIRIIPHHNYNAAINKYNHDIALLELDEPLVLNSYVTPICI ADKEYTNIFLKFGSGYVSGWGRVFHKGRSALVLQYLRVPLVDRATCLRSTKFTIYNNMFCAGFHEGGRDSCQGDSGGPHVTE VEGTSFLTGIISWGEECAMKGKYGIYTKVSRYVNWIKEKTKLT
275 (SOD1)	MATKAVCVLKGDGPVQGIINFEQKESNGPVKVWGSIKGLTEGLHGFHVHEFGDNTAGCTSAGPHFNPLSRKHGGPKDEERHV GDLGNVTADKDGVADVSIEDSVISLSGDHCIIGRTLVVHEKADDLGKGGNEESTKTGNAGSRLACGVIGIAQ
276 (C9)	MSACRSFAVAICILEISILTAQYTTSYDPELTESSGSASHIDCRMSPWSEWSQCDPCLRQMFRSRSIEVFGQFNGKRCTDAVGDRR QCVPTEPCEDAEDDCGNDFQCSTGRCIKMRLRCNGDNDCGDFSDEDDCESEPRPPCRDRVVEESELARTAGYGINILGMDPLST PFDNEFYNGLCNRDRDGNTLTYYRRPWNVASLIYETKGEKNFRTEHYEEQIEAFKSIIQEKTSNFNAAISLKFTPTETNKAEQCC EETASSISLHGKGSFRFSYSKNETYQLFLSYSSKKEKMFLHVKGEIHLGRFVMRNRDVVLTTTFVDDIKALPTTYEKGEYFAFLE TYGTHYSSSGSLGGLYELIYVLDKASMKRKGVELKDIKRCLGYHLDVSLAFSEISVGAEFNKDDCVKRGEGRAVNITSENLIDD VVSLIRGGTRKYAFELKEKLLRGTVIDVTDFVNWASSINDAPVLISQKLSPIYNLVPVKMKNAHLKKQNLERAIEDYINEFSVRK CHTCQNGGTVILMDGKCLCACPFKFEGIACEISKQKISEGLPALEFPNEK
277 (HTT)	MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQPPPPPPPPPPQLPQPPPQAQPLLPQPQPPPPPPPPGPAVAEEPL HRPKKELSATKKDRVNHCLTICENIVAQSVRNSPEFQKLLGIAMELFLLCSDDAESDVRMVADECLNKVIKALMDSNLPRLQLE LYKEIKKNGAPRSLRAALWRFAELAHLVRPQKCRPYLVNLLPCLTRTSKRPEESVQETLAAAVPKIMASFGNFANDNEIKVLLK AFIANLKSSSPTIRRTAAGSAVSICQHSRRTQYFYSWLLNVLLGLLVPVEDEHSTLLILGVLLTLRYLVPLLQQVKDTSLKGSFG VTRKEMEVSPSAEQLVQVYELTLHHTQHDHNVVTGALELLQQLFRTPPPELLQTLTAVGGIGQLTAAKEESGGRSRSGSIVELI AGGGSSCSPVLSRKQKGKVLLGEEEALEDDSESRSDVSSSALTASVKDEISGELAASSGVSTPGSAGHDIITEQPRSQHTLQADSV DLASCDLTSSATDGDEEDILSHSSSQVSAVPSDPAMDLNDGTQASSPISDSSQTTTEGPDSAVTPSDSSEIVLDGTDNQYLGLQIG QPQDEDEEATGILPDEASEAFRNSSMALQQAHLLKNMSHCRQPSDSSVDKFVLRDEATEPGDQENKPCRIKGDIGQSTDDDSAP LVHCVRLLSASFLLTGGKNVLVPDRDVRVSVKALALSCVGAAVALHPESFFSKLYKVPLDTTEYPEEQYVSDILNYIDHGDPQV RGATAILCGTLICSILSRSRFHVGDWMGTIRTLTGNTFSLADCIPLLRKTLKDESSVTCKLACTAVRNCVMSLCSSSYSELGLQLII DVLTLRNSSYWLVRTELLETLAEIDFRLVSFLEAKAENLHRGAHHYTGLLKLQERVLNNVVIHLLGDEDPRVRHVAAASLIRLV PKLFYKCDQGQADPVVAVARDQSSVYLKLLMHETQPPSHFSVSTITRIYRGYNLLPSITDVTMENNLSRVIAAVSHELITSTTRA LTFGCCEALCLLSTAFPVCIWSLGWHCGVPPLSASDESRKSCTVGMATMILTLLSSAWFPLDLSAHQDALILAGNLLAASAPKSL RSSWASEEEANPAATKQEEVWPALGDRALVPMVEQLFSHLLKVINICAHVLDDVAPGPAIKAALPSLTNPPSLSPIRRKGKEKEP GEQASVPLSPKKGSEASAASRQSDTSGPVTTSKSSSLGSFYHLPSYLKLHDVLKATHANYKVTLDLQNSTEKFGGFLRSALDVL

SEQ ID NO: (Protein)	Sequence
	SQILELATLQDIGKCVEEILGYLKSCFSREPMMATVCVQQLLKTLFGTNLASQFDGLSSNPSKSQGRAQRLGSSSVRPGLYHYCF MAPYTHFTQALADASLRNMVQAEQENDTSGWFDVLQKVSTQLKTNLTSVTKNRADKNAIHNHIRLFEPLVIKALKQYTTTTC VQLQKQVLDLLAQLVQLRVNYCLLDSDQVFIGFVLKQFEYIEVGQFRESEAIIPNIFFFLVLLSYERYHSKQIIGIPKIIQLCDGIMA SGRKAVTHAIPALQPIVHDLFVLRGTNKADAGKELETQKEVVVSMLLRLIQYHQVLEMFILVLQQCHKENEDKWKRLSRQIADI ILPMLAKQQMHIDSHEALGVLNTLFEILAPSSLRPVDMLLRSMFVTPNTMASVSTVQLWISGILAILRVLISQSTEDIVLSRIQELS FSPYLISCTVINRLRDGDSTSTLEEHSEGKQIKNLPEETFSRFLLQLVGILLEDIVTKQLKVEMSEQQHTFYCQELGTLLMCLIHIF KSGMFRRITAAATRLFRSDGCGGSFYTLDSLNLRARSMITTHPALVLLWCQILLLVNHTDYRWWAEVQQTPKRHSLSSTKLLSP QMSGEEEDSDLAAKLGMCNREIVRRGALILFCDYVCQNLHDSEHLTWLIVNHIQDLISLSHEPPVQDFISAVHRNSAASGLFIQAI QSRCENLSTPTMLKKTLQCLEGIHLSQSGAVLTLYVDRLLCTPFRVLARMVDILACRRVEMLLAANLQSSMAQLPMEELNRIQE YLQSSGLAQRHQRLYSLLDRFRLSTMQDSLSPSPPVSSHPLDGDGHVSLETVSPDKDWYVHLVKSQCWTRSDSALLEGAELVN RIPAEDMNAFMMNSEFNLSLLAPCLSLGMSEISGGQKSALFEAAREVTLARVSGTVQQLPAVHHVFQPELPAEPAAYWSKLND LFGDAALYQSLPTLARALAQYLVVVSKLPSHLHLPPEKEKDIVKFVVATLEALSWHLIHEQIPLSLDLQAGLDCCCLALQLPGL WSVVSSTEFVTHACSLIYCVHFILEAVAVQPGEQLLSPERRTNTPKAISEEEEVDPNTQNPKYITAACEMVAEMVESLQSVLAL GHKRNSGVPAFLTPLLRNIIISLARLPLVNSYTRVPPLVWKLGWSPKPGGDFGTAFPEIPVEFLQEKEVFKEFIYRINTLGWTSRT QFEETWATLLGVLVTQPLVMEQEESPPEEDTERTQINVLAVQAITSLVLSAMTVPVAGNPAVSCLEQQPRNKPLKALDTRFGRK LSIIRGIVEQEIQAMVSKRENIATHHLYQAWDPVPSLSPATTGALISHEKLLLQINPERELGSMSYKLGQVSIHSVWLGNSITPLRE EEWDEEEEEADAPAPSSPPTSPVNSRKHRAGVDIHSCSQFLLELYSRWILPSSSARRTPAILISEVVRSLLVVSDLFTERNQFELM YVTLTELRRVHPSEDEILAQYLVPATCKAAAVLGMDKAVAEPVSRLLESTLRSSHLPSRVGALHGVLYVLECDLLDDTAKQLIP VISDYLLSNLKGIAHCVNIHSQQHVLVMCATAFYLIENYPLDVGPEFSASIIQMCGVMLSGSEESTPSIIYHCALRGLERLLLSEQL SRLDAESLVKLSVDRVNVHSPHRAMAALGLMLTCMYTGKEKVSPGRTSDPNPAAPDSESVIVAMERVSVLFDRIRKGFPCEAR VVARILPQFLDDFFPPQDIMNKVIGEFLSNQQPYPQFMATVVYKVFQTLHSTGQSSMVRDWVMLSLSNFTQRAPVAMATWSLS CFFVSASTSPWVAAILPHVISRMGKLEQVDVNLFCLVATDFYRHQIEEELDRRAFQSVLEVVAAPGSPYHRLLTCLRNVHKVTT C
278 (MECP2)	MVAGMLGLREEKSEDQDLQGLKDKPLKFKKVKKDKKEEKEGKHEPVQPSAHHSAEPAEAGKAETSEGSGSAPAVPEASASPK QRRSIIRDRGPMYDDPTLPEGWTRKLKQRKSGRSAGKYDVYLINPQGKAFRSKVELIAYFEKVGDTSLDPNDFDFTVTGRGSPS RREQKPPKKPKPSPKAPGTGRGRGRPKGSGTTRPKAATSEGVQVKRVLEKSPGKLLVKMPFQTSPGGKAEGGGATTSTQVMVIK RPGRKRKAEADPQAIPKKRGRKPGSVVAAAAAEAKKAVKESSIRSVQETVLPIKKRKTRETVSIEVKEVVKPLLVSTLGEKSG KGLKTCKSPGRKSKESSPKGRSSSASSPPKKEHHHHHHHHSESPKAPVPLLPPLPPPPPEPESSEDPTSPPEPQDLSSSVCKEEKMPR GGSLESDGCPKEPAKTQPAVATAATAAEKYKHRGEGERKDIVSSSMPRPNREEPVDSRTPVTERVS
279 (SMN1)	MAMSSGGSGGGVPEQEDSVLFRRGTGQSDDSDIWDDTALIKAYDKAVASFKHALKNGDICETSGKPKTTPKRKPAKKNKSQK KNTAASLQQWKVGDKCSAIWSEDGCIYPATIASIDFKRETCVVVYTGYGNREEQNLSDLLSPICEVANNIEQNAQENENESQVS

SEQ ID NO: (Protein)	Sequence
	TDESENSRSPGNKSDNIKPKSAPWNSFLPPPPPMPGPRLGPGKPGLKFNGPPPPPPPPPPHLLSCWLPPFPSGPPIIPPPPPICPDSLD DADALGSMLISWYMSGYHTGYYMGFRQNQKEGRCSHSLN
280 (TARDBP)	MSEYIRVTEDENDEPIEIPSEDDGTVLLSTVTAQFPGACGLRYRNPVSQCMRGVRLVEGILHAPDAGWGNLVYVVNYPKDNKR KMDETDASSAVKVKRAVQKTSDLIVLGLPWKTTEQDLKEYFSTFGEVLMVQVKKDLKTGHSKGFGFVRFTEYETQVKVMSQ RHMIDGRWCDCKLPNSKQSQDEPLRSRKVFVGRCTEDMTEDELREFFSQYGDVMDVFIPKPFRAFAFVTFADDQIAQSLCGED LIIKGISVHISNAEPKHNSNRQLERSGRFGGNPGGFGNSRGGGAGLGNNQGSNMGGGMNFGAFSINPAMMAAAQAA LQSSWGMMGMLASQQNQSGPSGNNQNQGNMQREPNQAFGSGNNSYSGSNSGAAIGWGSASNAGSGSGFNGGFGSSMDSKSS GWGM
281 (FUS)	MASNDYTQQATQSYGAYPTQPGQGYSQQSSQPYGQQSYSGYSQSTDTSGYGQSSYSSYGQSQNTGYGTQSTPQGYGSTGGYG SSQSSQSSYGQQSSYPGYGQQPAPSSTSGSYGSSSQSSSYGQPQSGSYSQQPSYGGQQQSYGQQQSYNPPQGYGQQNQYNSSSG GGGGGGGGNYGQDQSSMSSGGGSGGYGNQDQSGGGGSGGYGQQDRGGRGRGGSGGGGGGGYNRSSGGYEPRGRG GGRGGRGGMGGSDRGGFNKFGGPRDQGSRHDSEQDNSDNNTIFVQGLGENVTIESVADYFKQIGIIKTNKKTGQPMINLYTDR ETGKLKGEATVSFDDPPSAKAAIDWFDGKEFSGNPIKVSFATRRADFNRGGGNGRGGRGRGGPMGRGGYGGGGSGGGGRGGF PSGGGGGGGQQRAGDWKCPNPTCENMNFSWRNECNQCKAPKPDGPGGGPGGSHMGGNYGDDRRGGRGGYDRGGYRGRGG DRGGFRGGRGGGDRGGFGPGKMDSRGEHRQDRRERPY
282 (USH2A)	MNCPVLSLGSGFLFQVIEMLIFAYFASISLTESRGLFPRLENVGAFKKVSIVPTQAVCGLPDRSTFCHSSAAAESIQFCTQRFCIQD CPYRSSHPTYTALFSAGLSSCITPDKNDLHPNAHSNSASFIFGNHKSCFSSPPSPKLMASFTLAVWLKPEQQGVMCVIEKTVDGQI VFKLTISEKETMFYYRTVNGLQPPIKVMTLGRILVKKWIHLSVQVHQTKISFFINGVEKDHTPFNARTLSGSITDFASGTVQIGQS LNGLEQFVGRMQDFRLYQVALTNREILEVFSGDLLRLHAQSHCRCPGSHPRVHPLAQRYCIPNDAGDTADNRVSRLNPEAHPL SFVNDNDVGTSWVSNVFTNITQLNQGVTISVDLENGQYQVFYIIIIQFFSPQPTEIRIQRKKENSLDWEDWQYFARNCGAFGMKN NGDLEKPDSVNCLQLSNFTPYSRGNVTFSILTPGPNYRPGYNNFYNTPSLQEFVKATQIRFHFHGQYYTTETAVNLRHRYYAVD EITISGRCQCHGHADNCDTTSQPYRCLCSQESFTEGLHCDRCLPLYNDKPFRQGDQVYAFNCKPCQCNSHSKSCHYNISVDPFPF EHFRGGGVCDDCEHNTTGRNCELCKDYFFRQVGADPSAIDVCKPCDCDTVGTRNGSILCDQIGGQCNCKRHVSGRQCNQCQ NGFYNLQELDPDGCSPCNCNTSGTVDGDITCHQNSGQCKCKANVIGLRCDHCNFGFKFLRSFNDVGCEPCQCNLHGSVNKFCN PHSGQCECKKEAKGLQCDTCRENFYGLDVTNCKACDCDTAGSLPGTVCNAKTGQCICKPNVEGRQCNKCLEGNFYLLRQNNSF LCLPCNCDKTGTINGSLLCNKSTGQCPCKLGVTGLRCNQCEPHRYNLTIDNFQHCQMCECDSLGTLPGTICDPISGQCLCVPNRQ GRRCNQCQPGFYISPGNATGCLPCSCHTTGAVNHICNSLTGQCVCQDASIAGQRCDQCKDHYFGFDPQTGRCQPCNCHLSGAL NETCHLVTGQCFCKQFVTGSKCDACVPSASHLDVNNLLGCSKTPFQQPPPRGQVQSSSAINLSWSPPDSPNAHWLTYSLLRDGF EIYTTEDQYPYSIQYFLDTDLLPYTKYSYYIETTNVHGSTRSVAVTYKTKPGVPEGNLTLSYIIPIGSDSVTLTWTTLSNQSGPIEK YILSCAPLAGGQPCVSYEGHETSATIWNLVPFAKYDFSVQACTSGGCLHSLPITVTTAQAPPQRLSPPKMQKISSTELHVEWSPP

SEQ ID NO: (Protein)	Sequence
	AELNGIIIRYELYMRRLRSTKETTSEESRVFQSSGWLSPHSFVESANENALKPPQTMTTITGLEPYTKYEFRVLAVNMAGSVSSA
	WVSERTGESAPVFMIPPSVFPLSSYSLNISWEKPADNVTRGKVVGYDINMLSEQSPQQSIPMAFSQLLHTAKSQELSYTVEGLKP
	YRIYEFTITLCNSVGCVTSASGAGQTLAAAPAQLRPPLVKGINSTTIHLRWFPPEELNGPSPIYQLERRESSLPALMTTMMKGIRFI
	GNGYCKFPSSTHPVNTDFTGIKASFRTKVPEGLIVFAASPGNQEEYFALQLKKGRLYFLFDPQGSPVEVTTTNDHGKQYSDGKW
	HEIIAIRHQAFGQITLDGIYTGSSAILNGSTVIGDNTGVFLGGLPRSYTILRKDPEIIQKGFVGCLKDVHFMKNYNPSAIWEPLDWQ
	SSEEQINVYNSWEGCPASLNEGAQFLGAGFLELHPYMFHGGMNFEISFKFRTDQLNGLLLFVYNKDGPDFLAMELKSGILTFRL
	NTSLAFTQVDLLLLGLSYCNGKWNKVIIKKEGSFISASVNGLMKHASESGDQPLVVNSPVYVGGIPQELLNSYQHLCLEQGFGGC
	MKDVKFTRGAVVNLASVSSGAVRVNLDGCLSTDSAVNCRGNDSILVYQGKEQSVYEGGLQPFTEYLYRVIASHEGGSVYSDW
	SRGRTTGAAPQSVPTPSRVRSLNGYSIEVTWDEPVVRGVIEKYILKAYSEDSTRPPRMPSASAEFVNTSNLTGILTGLLPFKNYAV
	TLTACTLAGCTESSHALNISTPQEAPQEVQPPVAKSLPSSLLLSWNPPKKANGIITQYCLYMDGRLIYSGSEENYIVTDLAVFTPH
	QFLLSACTHVGCTNSSWVLLYTAQLPPEHVDSPVLTVLDSRTIHIQWKQPRKISGILERYVLYMSNHTHDFTIWSVIYNSTELFQ
	DHMLQYVLPGNKYLIKLGACTGGGCTVSEASEALTDEDIPEGVPAPKAHSYSPDSFNVSWTEPEYPNGVITSYGLYLDGILIHNS
	SELSYRAYGFAPWSLHSFRVQACTAKGCALGPLVENRTLEAPPEGTVNVFVKTQGSRKAHVRWEAPFRPNGLLTHSVLFTGIF
	YVDPVGNNYTLLNVTKVMYSGEETNLWVLIDGLVPFTNYTVQVNISNSQGSLITDPITIAMPPGAPDGVLPPRLSSATPTSLQVV
	WSTPARNNAPGSPRYQLQMRSGDSTHGFLELFSNPSASLSYEVSDLQPYTEYMFRLVASNGFGSAHSSWIPFMTAEDKPGPVVP
	PILLDVKSRMMLVTWQHPRKSNGVITHYNIYLHGRLYLRTPGNVTNCTVMHLHPYTAYKFQVEACTSKGCSLSPESQTVWTLP
	GAPEGIPSPELFSDTPTSVIISWQPPTHPNGLVENFTIERRVKGKEEVTTLVTLPRSHSMRFIDKTSALSPWTKYEYRVLMSTLHG
	GTNSSAWVEVTTRPSRPAGVQPPVVTVLEPDAVQVTWKPPLIQNGDILSYEIHMPDPHITLTNVTSAVLSQKVTHLIPFTNYSVTI
	VACSGGNGYLGGCTESLPTYVTTHPTVPQNVGPLSVIPLSESYVVISWQPPSKPNGPNLRYELLRRKIQQPLASNPPEDLNRWHN
	IYSGTQWLYEDKGLSRFTTYEYMLFVHNSVGFTPSREVTVTTLAGLPERGANLTASVLNHTAIDVRWAKPTVQDLQGEVEYYT
	LFWSSATSNDSLKILPDVNSHVIGHLKPNTEYWIFISVFNGVHSINSAGLHATTCDGEPQGMLPPEVVIINSTAVRVIWTSPSNPN
	GVVTEYSIYVNNKLYKTGMNVPGSFILRDLSPFTIYDIQVEVCTIYACVKSNGTQITTVEDTPSDIPTPTIRGITSRSLQIDWVSPRK
	PNGIILGYDLLWKTWYPCAKTQKLVQDQSDELCKAVRCQKPESICGHICYSSEAKVCCNGVLYNPKPGHRCCEEKYIPFVLNST
	GVCCGGRIQEAQPNHQCCSGYYARILPGEVCCPDEQHNRVSVGIGDSCCGRMPYSTSGNQICCAGRLHDGHGQKCCGRQIVSN
	DLECCGGEEGVVYNRLPGMFCCGQDYVVNMSDTICCSASSGESKAHIKKNDPVPVKCCETELIPKSQKCCNGVGYNPLKYVCSD
	KISTGMMMKETKECRILCPASMEATEHCGRCDFNFTSHICTVIRGSHNSTGKASIEEMCSSAEETIHTGSVNTYSYTDVNLKPYM
	TYEYRISAWNSYGRGLSKAVRARTKEDVPQGVSPPTWTKIDNLEDTIVLNWRKPIQSNGPIIYYILLRNGIERFRGTSLSFSDKEG
	IQPFQEYSYQLKACTVAGCATSSKVVAATTQGVPESILPPSITALSAVALHLSWSVPEKSNGVIKEYQIRQVGKGLIHTDTTDRR
	QHTVTGLQPYTNYSFTLTACTSAGCTSSEPFLGQTLQAAPEGVWVTPRHIINSTTVELYWSLPEKPNGLVSQYQLSRNGNLLFL
	GGSEEQNFTDKNLEPNSRYTYKLEVKTGGGSSASDDYIVQTPMSTPEEIYPPYNITVIGPYSIFVAWIPPGILIPEIPVEYNVLLNDG
	SVTPLAFSVGHHQSTLLENLTPFTQYEIRIQACQNGSCGVSSRMFVKTPEAAPMDLNSPVLKALGSACIEIKWMPPEKPNGIIINY
	FIYRRPAGIEEESVLFVWSEGALEFMDEGDTLRPFTLYEYRVRACNSKGSVESLWSLTQTLEAPPQDFPAPWAQATSAHSVLLN
	WTKPESPNGIISHYRVVVYQERPDDPTFNSPTVHAFTVKGTSHQAHLYGLEPFTTYRIGVVAANHAGEILSPWTLIQTLESSPSGLR

SEQ ID NO: (Protein)	Sequence
	NFIVEQKENGRALLLQWSEPMRTNGVIKTYNIFSDGFLEYSGLNRQFLFRRLDPFTLYTLTLEACTRAGCAHSAPQPLWTDEAPP DSQLAPTVHSVKSTSVELSWSEPVNPNGKIIRYEVIRRCFEGKAWGNQTIQADEKIVFTEYNTERNTFMYNDTGLQPWTQCEYK IYTWNSAGHTCSSWNVVRTLQAPPEGLSPPVISYVSMNPQKLLISWIPPEQSNGIIQSYRLQRNEMLYPFSFDPVTFNYTDEELLP FSTYSYALQACTSGGCSTSKPTSITTLEAAPSEVSPPDLWAVSATQMNVCWSPPTVQNGKITKYLVRYDNKESLAGQGLCLLVS HLQPYSQYNFSLVACTNGGCTASVSKSAWTMEALPENMDSPTLQVTGSESIEITWKPPRNPNGQIRSYELRRDGTIVYTGLETR YRDFTLTPGVEYSYTVTASNSQGGILSPLVKDRTSPSAPSGMEPPKLQARGPQEILVNWDPPVRTNGDIINYTLFIRELFERETKII HINTTHNSFGMQSYIVNQLKPFHRYEIRIQACTTLGCASSDWTFIQTPEIAPLMQPPPHLEVQMAPGGFQPTVSLLWTGPLQPNG KVLYYELYRRQIATQPRKSNPVLIYNGSSTSFIDSELLPFTEYEYQVWAVNSAGKAPSSWTWCRTGPAPPEGLRAPTFHVISSTQ AVVNISAPGKPNGIVSLYRLFSSSAHGAETVLSEGMATQQTLHGLQAFTNYSIGVEACTCFNCCSKGPTAELRTHPAPPSGLSSP QIGTLASRTASFRWSPPMFPNGVIHSYELQFHVACPPDSALPCTPSQIETKYTGLGQKASLGGLQPYTTYKLRVVAHNEVGSTAS EWISFTTQKELPQYRAPFSVDSNLSVVCVNWSDTFLLNGQLKEYVLTDGGRRVYSGLDTTLYIPRTADKTFFFQVICTTDEGSV KTPLIQYDTSTGLGLVLTTPGKKKGSRSKSTEFYSELWFIVLMAMLGLILLAIFLSLILQRKIHKEPYIRERPPLVPLQKRMSPLNV YPPGENHMGLADTKIPRSGTPVSIRSNRSACVLRIPSQNQTSLTYSQGSLHRSVSQLMDIQDKKVLMDNSLWEAIMGHNSGLYV DEEDLMNAIKDFSSVTKERTTFTDTHL
1824 (A1AT)	MPSSVSWGILLLAGLCCLVPVSLAEDPQGDAAQKTDTSHHDQDHPTFNKITPNLAEFAFSLYRQLAHQSNSTNIFFSPVSIATAF AMLSLGTKADTHDEILEGLNFNLTEIPEAQIHEGFQELLRTLNQPDSQLQLTTGNGLFLSEGLKLVDKFLEDVKKLYHSEAFTVN FGDTEEAKKQINDYVEKGTQGKIVDLVKELDRDTVFALVNYIFFKGKWERPFEVKDTEEEDFHVDQVTTVKVPMMKRLGMFN IQHCKKLSSWVLLMKYLGNATAIFFLPDEGKLQHLENELTHDIITKFLENEDRRSASLHLPKLSITGTYDLKSVLGQLGITKVFSN GADLSGVTEEAPLKLSKAVHKAVLTIDEKGTEAAGAMFLEAIPMSIPPEVKFNKPFVFLMIEQNTKSPLFMGKVVNPTQK

[688] **TABLE 12** provides exemplary disease and syndromes that may be treated, prevented, or inhibited in a subject using the compositions and methods described herein.

TABLE 12. Exemplary Diseases and Syndromes

11-hydroxylase deficiency; 17,20-desmolase deficiency; 17-hydroxylase deficiency; 3-hydroxyisobutyrate aciduria; 3-hydroxysteroid dehydrogenase deficiency; 46,XY gonadal dysgenesis; AAA syndrome; ABCA3 deficiency; ABCC8-associated hyperinsulinism; aceruloplasminemia; acromegaly; achondrogenesis type 2; acral peeling skin syndrome; acrodermatitis enteropathica; acute bacterial infection; adrenocortical micronodular hyperplasia;

adrenoleukodystrophies; adrenomyeloneuropathies; Aicardi-Goutieres syndrome; AIDS; Alagille disease (also called Alagille Syndrome); Alexander Disease; Alpers syndrome; alpha-1 antitrypsin deficiency (AATD); alpha-mannosidosis; Alstrom syndrome; Alzheimer's disease; amebic dysentery; amelogenesis imperfecta; amish type microcephaly; amyotrophic lateral sclerosis (ALS); anaplastic large cell lymphoma; anauxetic dysplasia; androgen insensitivity syndrome; angiopathic thrombosis; antiphospholipid syndrome; Antley-Bixler syndrome; APECED; Apert syndrome; aplasia of lacrimal and salivary glands; arginase-1 deficiency; argininosuccinic aciduria; argininemia; arrhythmogenic right ventricular dysplasia; Arts syndrome; ARVD2; arylsulfatase deficiency type metachromatic leukodystrophy; ataxia telangiectasia; atherosclerotic cardiovascular disease; autoimmune lymphoproliferative syndrome; autoimmune polyglandular syndrome type 1; autosomal dominant anhidrotic ectodermal dysplasia; autosomal dominant deafness; autosomal dominant polycystic kidney disease; autosomal recessive microtia; autosomal recessive renal glucosuria; autosomal visceral heterotaxy; babesiosis; bacterial vaginosis; balantidial dysentery; Bardet-Biedl syndrome; Bartter syndrome; basal cell nevus syndrome; Batten disease; benign recurrent intrahepatic cholestasis; beta-mannosidosis; β-thalassemia; Bethlem myopathy; Blackfan-Diamond anemia; bleeding disorder (coagulation); blepharophimosis; Byler disease; C syndrome; CADASIL; calcific aortic stenosis; calcification of joints and arteries; carbamoyl phosphate synthetase I deficiency; carcinoid syndrome diarrhea; cardiofaciocutaneous syndrome; cardiovascular disease (CVD); Carney triad; carnitine palmitoyltransferase deficiencies; cartilage-hair hypoplasia; cblC type of combined methylmalonic aciduria; CD18 deficiency; CD3Z-associated primary T-cell immunodeficiency; CD40L deficiency; CDAGS syndrome; CDG1A; CDG1B; CDG1M; CDG2C; CEDNIK syndrome; central core disease; centronuclear myopathy; cerebral capillary malformation; cerebrooculofacioskeletal syndrome type 4; cerebrooculogacioskeletal syndrome; cerebrotendinous xanthomatosis; Chaga's Disease; Charcot Marie Tooth Disesase; chemotherapy; cherubism; CHILD syndrome; chronic granulomatous disease; chronic recurrent multifocal osteomyelitis; cirrhosis; citrin deficiency; citrullinemia type I; citrullinemia type II; classic hemochromatosis; CNPPB syndrome; cobalamin C disease; Cockayne syndrome; coenzyme Q10 deficiency; Coffin-Lowry syndrome; Cohen syndrome; combined deficiency of coagulation factors V; common variable immune deficiency 3; complement hyperactivation; complete androgen insentivity; cone rod dystrophies; conformational diseases; congenital adrenal hyperplasia; congenital bile adid synthesis defect type 1; congenital bile adid synthesis defect type 2; congenital defect in bile acid synthesis type; congenital erythropoietic porphyria; congenital generalized osteosclerosis; congenital hyperplasia (CAH); congenital muscular dystrophy 1A (MDC1A); Cornelia de Lange syndrome; coronary heart disease; Cousin syndrome; Cowden disease; COX deficiency; Cri du chat syndrome; Crigler-Najjar disease; Crigler-Najjar syndrome type 1; Crisponi syndrome; Crouzon syndrome; Currarino syndrome; Curth-Macklin type ichthyosis hystrix; cutaneous T-cell lymphoma; cutis laxa; cystic fibrosis; cystinosis;

d-2-hydroxyglutaric aciduria; DDP syndrome; Dejerine-Sottas disease; Denys-Drash syndrome; Dercum disease; desmin cardiomyopathy; desmin myopathy; DGUOK-associated mitochondrial DNA depletion; diabetes Type I; diabetes Type II; disorders of glutamate metabolism; distal spinal muscular atrophy type 5; DNA repair diseases; dominant optic atrophy; Doyne honeycomb retinal dystrophy; Dravet Syndrome; Duchenne muscular dystrophy; dyskeratosis congenita; Ehlers-Danlos syndrome type 4; Ehlers-Danlos syndromes; Elejalde disease; Ellis-van Creveld disease; Emery-Dreifuss muscular dystrophies; encephalomyopathic mtDNA depletion syndrome; encephalitis; enzymatic diseases; EPCAM-associated congenital tufting enteropathy; epidermolysis bullosa with pyloric atresia; epilepsy; fabry disease; facioscapulohumeral muscular dystrophy; Factor V Leiden thrombophilia; Faisalabad histiocytosis; familial atypical mycobacteriosis; familial capillary malformation-arteriovenous; Familial Creutzfeld-Jakob disease; familial esophageal achalasia; familial glomuvenous malformation; familial hemophagocytic lymphohistiocytosis; familial mediterranean fever; familial megacalyces; familial schwannomatosis; familial spina bifida; familial splenic asplenia/hypoplasia; familial thrombotic thrombocytopenic purpura; Fanconi disease (Fanconi anemia); Feingold syndrome; FENIB; fibrodysplasia ossificans progressiva; FKTN; Fragile X syndrome; Francois-Neetens fleck corneal dystrophy; Frasier syndrome; Friedreich's ataxia; FTDP-17; Fuchs corneal dystrophy; fucosidosis; G6PD deficiency; galactosialidosis; Galloway syndrome; Gardner syndrome; Gaucher disease; Gitelman syndrome; glaucoma; GLUT1 deficiency; GM2- Gangliosidoses (*e.g.,* Tay Sachs Disease, Sandhoff Disease) glycogen storage disease type 1b; glycogen storage disease type 2; glycogen storage disease type 3; glycogen storage disease type 4; glycogen storage disease type 9a; glycogen storage diseases; GM1-gangliosidosis; Greenberg syndrome; Greig cephalopolysyndactyly syndrome; hair genetic diseases; hairy cell leukemia; HANAC syndrome; harlequin type ichtyosis congenita; HDR syndrome; hearing loss; heart failure; hemochromatosis type 3; hemochromatosis type 4; hemolytic anemia; hemolytic uremic syndrome; hemophilia A; hemophilia B; hepatitis C infection; hereditary angioedema type 3; hereditary angioedemas; hereditary hemorrhagic telangiectasia; hereditary hypofibrinogenemia; hereditary intraosseous vascular malformation; hereditary leiomyomatosis and renal cell cancer; hereditary neuralgic amyotrophy; hereditary orotic aciduria; hereditary sensory and autonomic neuropathy type; Hermansky-Pudlak disease; HHH syndrome; HHT2; hidrotic ectodermal dysplasia type 1; hidrotic ectodermal dysplasias; histiocytic sarcoma; HNF4A-associated hyperinsulinism; HNPCC; homozygous familial hypercholesterolemia; hormone refractory prostate cancer; human immunodeficiency with microcephaly; Human monkeypox (MPX); human papilloma virus (HPV) infection; Huntington's disease; hyper-IgD syndrome; hyperinsulinism-hyperammonemia syndrome; hypercholesterolemia; hypertension; hypertrophy of the retinal pigment epithelium; hypochondrogenesis; hypohidrotic ectodermal dysplasia; hypotension; ICF syndrome; idiopathic congenital intestinal pseudo-obstruction; immunodeficiency 13; immunodeficiency 17; immunodeficiency 25; immunodeficiency with hyper-IgM type 1;

immunodeficiency with hyper-IgM type 3; immunodeficiency with hyper-IgM type 4; immunodeficiency with hyper-IgM type 5; immunoglobulin alpha deficiency; inborn errors of thyroid metabolism; infantile myofibromatosis; infantile visceral myopathy; infantile X-linked spinal muscular atrophy; influenza A; influenza B; intradialytic hypotension; intrahepatic cholestasis of pregnancy; invasive aspergillosis; invasive mucormycosis; IPEX syndrome; IRAK4 deficiency; isolated congenital asplenia; Jeune syndrome; Johanson-Blizzard syndrome; Joubert syndrome; JP-HHT syndrome; juvenile hemochromatosis; juvenile hyalin fibromatosis; juvenile nephronophthisis; Kabuki mask syndrome; Kallmann syndromes; Kartagener syndrome; KCNJ11-associated hyperinsulinism; Kearns-Sayre syndrome; Kostmann disease; Kozlowski type of spondylometaphyseal dysplasia; Krabbe disease; LADD syndrome; late infantile-onset neuronal ceroid lipofuscinosis; LCK deficiency; LDHCP syndrome; Leber Congenital Amaurosis Teyp 10; Legius syndrome; Leigh syndrome; lethal congenital contracture syndrome 2; lethal congenital contracture syndromes; lethal contractural syndrome type 3; lethal neonatal CPT deficiency type 2; lethal osteosclerotic bone dysplasia; leukocyte adhesion deficiency; Li Fraumeni syndrome; LIG4 syndrome; limb girdle muscular dystrophies (LGMD1B, LGMD2A, LGMD2B); lipodystrophy; lissencephaly type 1; lissencephaly type 3; Loeys-Dietz syndrome; low phospholipid-associated cholelithiasis; Lynch Syndrome; lysinuric protein intolerance; a lysosomal storage disease (*e.g.,* Hunter syndrome, Hurler syndrome); macular dystrophy; Maffucci syndrome; Majeed syndrome; malaria; mannose-binding protein deficiency; mantle cell lymphoma; Marfan disease; Marshall syndrome; MASA syndrome; mastocytosis; MCAD deficiency; McCune-Albright syndrome; MCKD2; Meckel syndrome; MECP2 Duplication Syndrome; Meesmann corneal dystrophy; megacystis-microcolon-intestinal hypoperistalsis; megaloblastic anemia type 1; MEHMO; MELAS; Melnick-Needles syndrome; MEN2s; meningitis; Menkes disease; metachromatic leukodystrophies; methymalonic acidemia due to transcobalamin receptor defect; methylmalonic acidurias; methylvalonic aciduria; microcoria-congenital nephrosis syndrome; microvillous atrophy; migraine; mitochondrial neurogastrointestinal encephalomyopathy; monilethrix; monosomy X; mosaic trisomy 9 syndrome; Mowat-Wilson syndrome; mucolipidosis type 2; mucolipidosis type Ma; mucolipidosis type IV; mucopolysaccharidoses; mucopolysaccharidosis type 3A; mucopolysaccharidosis type 3C; mucopolysaccharidosis type 4B; multiminicore disease; multiple acyl-CoA dehydrogenation deficiency; multiple cutaneous and mucosal venous malformations; multiple endocrine neoplasia type 1; multiple myeloma; multiple sclerosis; multiple sulfatase deficiency; mycosis fungoides; myotonic dystrophy; NAIC; nail-patella syndrome; nemaline myopathies; neonatal diabetes mellitus; neonatal surfactant deficiency; nephronophtisis; Netherton disease; neurofibromatoses; neurofibromatosis type 1; Niemann-Pick disease type A; Niemann-Pick disease type B; Niemann-Pick disease type C; NKX2E; non-alcoholic fatty liver disease (NAFLD); non-alcoholic steatohepatitis (NASH); Noonan syndrome; North American Indian childhood cirrhosis; NROB1 duplication-associated DSD; ocular genetic

diseases; oculo-auricular syndrome; OLEDAID; oligomeganephronic renal hypolasia; Ollier disease; Opitz-Kaveggia syndrome; ornithine transcarbamylase deficiency (OTCD); orofaciodigital syndrome type 1; orofaciodigital syndrome type 2; osseous Paget disease; osteogenesis imperfecta; otopalatodigital syndrome type 2; orthostatic hypotension; overactive bladder; OXPHOS diseases; palmoplantar hyperkeratosis; panlobar nephroblastomatosis; Parkes-Weber syndrome; Parkinson's disease; partial deletion of 21q22.2-q22.3; Pearson syndrome; Pelizaeus-Merzbacher disease; Pendred syndrome; pentalogy of Cantrell; peroxisomal acyl-CoA-oxidase deficiency; Peutz-Jeghers syndrome; Pfeiffer syndrome; Pierson syndrome; pigmented nodular adrenocortical disease; pipecolic acidemia; Pitt-Hopkins syndrome; plasmalogens deficiency; platelet glycoprotein IV deficiency; pleuropulmonary blastoma and cystic nephroma; pneumonia; polycystic kidney disease; polycystic ovarian disease; polycystic lipomembranous osteodysplasia; Pompe disease; including infantile onset Pompe disease (IOPD) and late onset Pompe disease (LOPD); porphyrias; post-herpetic neuralgia; PRKAG2 cardiac syndrome; premature ovarian failure; primary erythermalgia; primary hemochromatoses; primary hyperoxaluria; progressive familial intrahepatic cholestasis; propionic acidemia; prostate cancer; protein-losing enteropathy; pulmonary arterial hypertension; pyruvate decarboxylase deficiency; RAPADILINO syndrome; renal cystinosis; restless leg syndrome; retinitis pigmentosa; Rett Syndrome; rhabdoid tumor predisposition syndrome; Rieger syndrome; ring chromosome 4; Roberts syndrome; Robinow-Sorauf syndrome; Rothmund-Thomson syndrome; severe combined immunodeficiency disorder (SCID); Saethre-Chotzen syndrome; Sandhoff disease; SC phocomelia syndrome; SCAS; Schinzel phocomelia syndrome; schizophrenia; severe hypertriglyceridemia; short rib-polydactyly syndrome type 1; short rib-polydactyly syndrome type 4; short-rib polydactyly syndrome type 2; short-rib polydactyly syndrome type 3; Shwachman disease; Shwachman-Diamond disease; sickle cell anemia; Silver-Russell syndrome; Simpson-Golabi-Behmel syndrome; skin infection; Smith-Lemli-Opitz syndrome; SPG7-associated hereditary spastic paraplegia; spherocytosis; spinocerebellar ataxia; spinal muscular atrophy; split-hand/foot malformation with long bone deficiencies; spondylocostal dysostosis; sporadic amyotrophic lateral sclerosis; sporadic visceral myopathy with inclusion bodies; storage diseases; Stargardt macular dystrophy; STRA6-associated syndrome; stroke; tardive dyskinesia; Tay-Sachs disease; thanatophoric dysplasia; thromboembolism; thrombosis; thrombophilia due to antithrombin III deficiency; thyroid metabolism diseases; Tourette syndrome; transcarbamylase deficiency; transthyretin-associated amyloidosis; trisomy 13; trisomy 22; trisomy 2p syndrome; tuberous sclerosis; tufting enteropathy; ullrich congenital muscular dystrophy (UCMD); urea cycle diseases; Usher Syndrome; Van Den Ende-Gupta syndrome; Van der Woude syndrome; variegated mosaic aneuploidy syndrome; VLCAD deficiency; von Hippel-Lindau disease; von Willebrand disease; Waardenburg syndrome; WAGR syndrome; Walker-Warburg syndrome; Werner syndrome; Wilson's disease; Wiskott-Aldrich Syndrome; Wolcott-Rallison syndrome; Wolfram

syndrome; X-linked agammaglobulinemia; X-linked chronic idiopathic intestinal pseudo-obstruction; X-linked cleft palate with ankyloglossia; X-linked dominant chondrodysplasia punctata; X-linked ectodermal dysplasia; X-linked Emery-Dreifuss muscular dystrophy; X-linked lissencephaly; X-linked lymphoproliferative disease; X-linked visceral heterotaxy; xanthinuria type 1; xanthinuria type 2; xeroderma pigmentosum; XPV; and Zellweger disease.

EXAMPLES

[689] The following examples are included for illustrative purposes only and are not intended to limit the scope of the invention.

Example 1: Identifying gRNAs targeting intron 1 of human ALB for effector protein CasPhi.12 or a variant thereof

[690] CasPhi.12 system comprising gRNA was tested for its indel activity within intron 1 of human albumin gene. Briefly, CasPhi.12 system included effector protein CasPhi.12 **(SEQ ID NO: 6)** or a variant thereof CasPhi.12 L26R **(SEQ ID NO: 228)**, and gRNA **(TABLE 13)** that were used for cleaving within intron 1 of human Albumin gene **(SEQ ID NO: 226)** in human liver carcinoma cell line Hep3B.

TABLE 13. Exemplary Compositions of Effector Protein and Guide Nucleic Acids

Comp NO:	Effector Protein SEQ ID NO:	gRNA ID NO:	PAM	gRNA Seq ID NO:	Comp NO:	Effector Protein SEQ ID NO:	gRNA ID NO:	PAM	gRNA Seq ID NO:
1	6 or 228	R9959	TTG	605	106	6 or 228	R10064	TTT	710
2	6 or 228	R9960	TTA	606	107	6 or 228	R10065	TTT	711
3	6 or 228	R9961	TTT	607	108	6 or 228	R10066	TTG	712
4	6 or 228	R9962	TTT	608	109	6 or 228	R10067	TTA	713
5	6 or 228	R9963	TTT	609	110	6 or 228	R10068	TTT	714
6	6 or 228	R9964	TTA	610	111	6 or 228	R10069	TTT	715
7	6 or 228	R9965	TTT	611	112	6 or 228	R10070	TTA	716
8	6 or 228	R9966	TTT	612	113	6 or 228	R10071	TTC	717
9	6 or 228	R9967	TTT	613	114	6 or 228	R10072	TTT	718
10	6 or 228	R9968	TTC	614	115	6 or 228	R10073	TTA	719
11	6 or 228	R9969	TTG	615	116	6 or 228	R10074	TTG	720
12	6 or 228	R9970	TTC	616	117	6 or 228	R10075	TTG	721
13	6 or 228	R9971	TTA	617	118	6 or 228	R10076	TTT	722
14	6 or 228	R9972	TTT	618	119	6 or 228	R10077	TTA	723
15	6 or 228	R9973	TTT	619	120	6 or 228	R10078	TTT	724
16	6 or 228	R9974	TTT	620	121	6 or 228	R10079	TTA	725
17	6 or 228	R9975	TTA	621	122	6 or 228	R10080	TTT	726
18	6 or 228	R9976	TTC	622	123	6 or 228	R10081	TTT	727
19	6 or 228	R9977	TTA	623	124	6 or 228	R10082	TTA	728
20	6 or 228	R9978	TTT	624	125	6 or 228	R10083	TTC	729
21	6 or 228	R9979	TTT	625	126	6 or 228	R10084	TTT	730
22	6 or 228	R9980	TTT	626	127	6 or 228	R10085	TTT	731
23	6 or 228	R9981	TTC	627	128	6 or 228	R10086	TTT	732
24	6 or 228	R9982	TTC	628	129	6 or 228	R10087	TTT	733

Comp NO:	Effector Protein SEQ ID NO:	gRNA ID NO:	PAM	gRNA Seq ID NO:
25	6 or 228	R9983	TTT	629
26	6 or 228	R9984	TTT	630
27	6 or 228	R9985	TTA	631
28	6 or 228	R9986	TTA	632
29	6 or 228	R9987	TTT	633
30	6 or 228	R9988	TTT	634
31	6 or 228	R9989	TTT	635
32	6 or 228	R9990	TTA	636
33	6 or 228	R9991	TTA	637
34	6 or 228	R9992	TTT	638
35	6 or 228	R9993	TTT	639
36	6 or 228	R9994	TTG	640
37	6 or 228	R9995	TTC	641
38	6 or 228	R9996	TTT	642
39	6 or 228	R9997	TTA	643
40	6 or 228	R9998	TTG	644
41	6 or 228	R9999	TTT	645
42	6 or 228	R10000	TTC	646
43	6 or 228	R10001	TTA	647
44	6 or 228	R10002	TTA	648
45	6 or 228	R10003	TTT	649
46	6 or 228	R10004	TTT	650
47	6 or 228	R10005	TTT	651
48	6 or 228	R10006	TTT	652
49	6 or 228	R10007	TTG	653
50	6 or 228	R10008	TTG	654
51	6 or 228	R10009	TTT	655
52	6 or 228	R10010	TTT	656
53	6 or 228	R10011	TTG	657
54	6 or 228	R10012	TTA	658
55	6 or 228	R10013	TTA	659
56	6 or 228	R10014	TTT	660
57	6 or 228	R10015	TTT	661
58	6 or 228	R10016	TTT	662
59	6 or 228	R10017	TTT	663
60	6 or 228	R10018	TTT	664
61	6 or 228	R10019	TTT	665
62	6 or 228	R10020	TTT	666
63	6 or 228	R10021	TTT	667

Comp NO:	Effector Protein SEQ ID NO:	gRNA ID NO:	PAM	gRNA Seq ID NO:
130	6 or 228	R10088	TTA	734
131	6 or 228	R10089	TTC	735
132	6 or 228	R10090	TTG	736
133	6 or 228	R10091	TTA	737
134	6 or 228	R10092	TTG	738
135	6 or 228	R10093	TTC	739
136	6 or 228	R10094	TTG	740
137	6 or 228	R10095	TTA	741
138	6 or 228	R10096	TTA	742
139	6 or 228	R10097	TTT	743
140	6 or 228	R10098	TTA	744
141	6 or 228	R10099	TTA	745
142	6 or 228	R10100	TTA	746
143	6 or 228	R10101	TTA	747
144	6 or 228	R10102	TTA	748
145	6 or 228	R10103	TTT	749
146	6 or 228	R10104	TTT	750
147	6 or 228	R10105	TTT	751
148	6 or 228	R10106	TTG	752
149	6 or 228	R10107	TTA	753
150	6 or 228	R10108	TTC	754
151	6 or 228	R10109	TTT	755
152	6 or 228	R10110	TTA	756
153	6 or 228	R10111	TTT	757
154	6 or 228	R10112	TTT	758
155	6 or 228	R10113	TTA	759
156	6 or 228	R10114	TTT	760
157	6 or 228	R10115	TTT	761
158	6 or 228	R10116	TTA	762
159	6 or 228	R10117	TTC	763
160	6 or 228	R10118	TTC	764
161	6 or 228	R10119	TTT	765
162	6 or 228	R10120	TTC	766
163	6 or 228	R10121	TTA	767
164	6 or 228	R10122	TTG	768
165	6 or 228	R10123	TTC	769
166	6 or 228	R10124	TTG	770
167	6 or 228	R10125	TTA	771
168	6 or 228	R10126	TTC	772

Comp NO:	Effector Protein SEQ ID NO:	gRNA ID NO:	PAM	gRNA Seq ID NO:
64	6 or 228	R10022	TTA	668
65	6 or 228	R10023	TTT	669
66	6 or 228	R10024	TTT	670
67	6 or 228	R10025	TTT	671
68	6 or 228	R10026	TTA	672
69	6 or 228	R10027	TTA	673
70	6 or 228	R10028	TTC	674
71	6 or 228	R10029	TTC	675
72	6 or 228	R10030	TTA	676
73	6 or 228	R10031	TTA	677
74	6 or 228	R10032	TTC	678
75	6 or 228	R10033	TTT	679
76	6 or 228	R10034	TTG	680
77	6 or 228	R10035	TTG	681
78	6 or 228	R10036	TTT	682
79	6 or 228	R10037	TTT	683
80	6 or 228	R10038	TTA	684
81	6 or 228	R10039	TTA	685
82	6 or 228	R10040	TTT	686
83	6 or 228	R10041	TTT	687
84	6 or 228	R10042	TTC	688
85	6 or 228	R10043	TTT	689
86	6 or 228	R10044	TTC	690
87	6 or 228	R10045	TTT	691
88	6 or 228	R10046	TTT	692
89	6 or 228	R10047	TTG	693
90	6 or 228	R10048	TTC	694
91	6 or 228	R10049	TTC	695
92	6 or 228	R10050	TTC	696
93	6 or 228	R10051	TTA	697
94	6 or 228	R10052	TTA	698
95	6 or 228	R10053	TTG	699
96	6 or 228	R10054	TTT	700
97	6 or 228	R10055	TTC	701
98	6 or 228	R10056	TTG	702
99	6 or 228	R10057	TTG	703
100	6 or 228	R10058	TTC	704
101	6 or 228	R10059	TTA	705
102	6 or 228	R10060	TTT	706

Comp NO:	Effector Protein SEQ ID NO:	gRNA ID NO:	PAM	gRNA Seq ID NO:
169	6 or 228	R10127	TTC	773
170	6 or 228	R10128	TTT	774
171	6 or 228	R10129	TTA	775
172	6 or 228	R10130	TTC	776
173	6 or 228	R10131	TTG	777
174	6 or 228	R10132	TTG	778
175	6 or 228	R10133	TTC	779
176	6 or 228	R10134	TTT	780
177	6 or 228	R10135	TTG	781
178	6 or 228	R10136	TTT	782
179	6 or 228	R10137	TTT	783
180	6 or 228	R10138	TTT	784
181	6 or 228	R10139	TTT	785
182	6 or 228	R10140	TTA	786
183	6 or 228	R10141	TTC	787
184	6 or 228	R10142	TTA	788
185	6 or 228	R10143	TTG	789
186	6 or 228	R10144	TTT	790
187	6 or 228	R10145	TTC	791
188	6 or 228	R10146	TTT	792
189	6 or 228	R10147	TTC	793
190	6 or 228	R10148	TTT	794
191	6 or 228	R10149	TTT	795
192	6 or 228	R10150	TTT	796
193	6 or 228	R10151	TTA	797
194	6 or 228	R10152	TTG	798
195	6 or 228	R10153	TTT	799
196	6 or 228	R10154	TTT	800
197	6 or 228	R10155	TTT	801
198	6 or 228	R10156	TTA	802
199	6 or 228	R10157	TTA	803
200	6 or 228	R10158	TTG	804
201	6 or 228	R10159	TTA	805
202	6 or 228	R10160	TTA	806
203	6 or 228	R10161	TTC	807
204	6 or 228	R10162	TTG	808
205	6 or 228	R10163	TTT	809
206	6 or 228	R10164	TTT	810
207	6 or 228	R10165	TTC	811

Comp NO:	Effector Protein SEQ ID NO:	gRNA ID NO:	PAM	gRNA Seq ID NO:
103	6 or 228	R10061	TTT	707
104	6 or 228	R10062	TTT	708
105	6 or 228	R10063	TTG	709

Comp NO:	Effector Protein SEQ ID NO:	gRNA ID NO:	PAM	gRNA Seq ID NO:
208	6 or 228	R10166	TTC	812
209	6 or 228	R10167	TTT	813
210	6 or 228	R10168	TTA	814

[691] For determining indel activity using compositions recited in **TABLE 13**, 20,000 Hep3B cells were seeded in 96-well plates. After 24 hours, 100 ng total RNA (1:1 mRNA to gRNA ratio) was transfected using Lipofectamine MessengerMax. 48 hours post-transfection, DNA was extracted and editing was quantified by amplicon sequencing.

[692] **FIGs. 1A-1F** illustrates indel activity for CasPhi.12 L26R mRNA transfected Hep3B cells.

[693] **FIG. 2** illustrates comparison of indel activity between CasPhi.12 L26R mRNA and CasPhi.12 mRNA transfected Hep3B cells.

[694] An analysis of **FIGs. 1A-1F** and **FIG. 2** indicate that CasPhi.12 system can be used for cleaving within intron 1 of human albumin gene.

Example 2: Identifying gRNAs targeting intron 1 of human ALB for effector protein CasPhi.32

[695] CasPhi.32 system was tested for its indel activity within intron 1 of human albumin gene. Briefly, CasPhi.32 system included an effector protein CasPhi.32 **(SEQ ID NO: 47)** and gRNA **(TABLE 14)** that were used for inserting a cut/cleavage within intron 1 of human albumin gene **(SEQ ID NO: 226)** in human liver carcinoma cell line Hep3B. Cas9 system was used as a positive control.

TABLE 14. Compositions of Effector Protein CasPhi.32 and Guide Nucleic Acids

Comp NO:	Effector Protein SEQ ID NO:	gRNA ID NO:	PAM	gRNA Seq ID NO:
1	47	R10169	GTTG	815
2	47	R10170	GTTC	816
3	47	R10171	GTTT	817
4	47	R10172	GTTT	818
5	47	R10173	GTTT	819
6	47	R10174	GTTA	820
7	47	R10175	GTTT	821
8	47	R10176	GTTA	822
9	47	R10177	GTTT	823
10	47	R10178	GTTA	824
11	47	R10179	GTTG	825
12	47	R10180	GTTT	826

Comp NO:	Effector Protein SEQ ID NO:	gRNA ID NO:	PAM	gRNA Seq ID NO:
14	47	R10182	GTTC	828
15	47	R10183	GTTG	829
16	47	R10184	GTTG	830
17	47	R10185	GTTG	831
18	47	R10186	GTTT	832
19	47	R10187	GTTC	833
20	47	R10188	GTTT	834
21	47	R10189	GTTT	835
22	47	R10190	GTTT	836
23	47	R10191	GTTA	837
24	47	R10192	GTTA	838
25	47	R10193	GTTT	839

Comp NO:	Effector Protein SEQ ID NO:	gRNA ID NO:	PAM	gRNA Seq ID NO:
13	47	R10181	GTTT	827

Comp NO:	Effector Protein SEQ ID NO:	gRNA ID NO:	PAM	gRNA Seq ID NO:

[696] For determining indel activity for CasPhi.32 system using compositions recited in **TABLE 14**, 20,000 Hep3B cells were seeded in 96-well plates. After 24 hours, cells were transfected with 100 ng total RNA (1:1 mRNA to gRNA ratio) using Lipofectamine MessengerMax. 48 hours post-transfection, DNA was extracted. DNA editing was quantified by amplicon sequencing.

[697] The results of amplicon sequencing are illustrated in **FIG. 3**.

[698] **FIG. 3** indicates that CasPhi.32 system can be used to insert a cut/cleavage within intron 1 of human albumin gene.

Example 3: Identifying gRNAs targeting intron 1 of human ALB for effector protein CasM.265466

[699] CasM.265466 system was tested for its indel activity within intron 1 of human albumin gene. Briefly, CasM.265466 system included the effector protein CasM.265466 **(SEQ ID NO: 231)** and gRNA **(TABLE 15)** that were used to insert a cut/cleavage within intron 1 of human albumin gene **(SEQ ID NO: 226)** in human liver carcinoma cell line Hep3B. Cas9 system was used as a positive control.

TABLE 15. Compositions of Effector Protein CasM.265466 and Guide Nucleic Acids

Comp NO:	Effector Protein SEQ ID NO:	gRNA ID NO:	PAM	gRNA Seq ID NO:
1	231	R10234	TCTG	840
2	231	R10235	TATG	841
3	231	R10236	TTTG	842
4	231	R10237	TTTG	843
5	231	R10238	TTTG	844
6	231	R10239	TTTG	845
7	231	R10240	TGTG	846
8	231	R10241	TCTG	847
9	231	R10242	TTTG	848
10	231	R10243	TTTG	849
11	231	R10244	TGTG	850
12	231	R10245	TCTG	851
13	231	R10246	TATG	852
14	231	R10247	TTTG	853

Comp NO:	Effector Protein SEQ ID NO:	gRNA ID NO:	PAM	gRNA Seq ID NO:
16	231	R10249	TTTG	855
17	231	R10250	TATG	856
18	231	R10251	TATG	857
19	231	R10252	TGTG	858
20	231	R10253	TCTG	859
21	231	R10254	TTTG	860
22	231	R10255	TCTG	861
23	231	R10256	TCTG	862
24	231	R10257	TCTG	863
25	231	R10258	TTTG	864
26	231	R10259	TATG	865
27	231	R10260	TCTG	866
28	231	R10261	TTTG	867
29	231	R10262	TTTG	868

15	231	R10248	TGTG	854					

[700] For determining indel activity for CasM.265466 system using compositions recited in **TABLE 15**, 20,000 Hep3B cells were seeded in 96-well plates. After 24 hours, cells were transfected with 100 ng total RNA (1:1 mRNA to gRNA ratio) using Lipofectamine MessengerMax. 48 hours post-transfection, DNA was extracted. DNA editing was quantified by amplicon sequencing.

[701] **FIG. 4** shows % indel activity of CasM.265466 system in combination with gRNA for cleaving a nucleotide sequence within intron 1 of human Albumin gene. An analysis of **FIG. 4** indicates that CasM.265466 system can be used to insert a nucleotide sequence encoding a donor nucleic acid within intron 1 of human Albumin gene with high percentage of total hits as well as high editing efficiency.

Example 4: Effector protein system for a donor nucleic acid in human cells

[702] An effector protein system is assayed for inserting a donor nucleic acid within a safe harbor loci. Briefly, the safe harbor loci is cleaved by the methods described in Examples 1-3, and a plasmid containing a donor nucleic acid is used for inserting the donor nucleic acid into a cleaved safe harbor loci. As described in Examples 1-3, the effector protein system includes an RNA that translates into an effector protein, and a gRNA. The effector protein comprises an amino acid sequence that is at least 90% identical to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**. The gRNA comprises a spacer sequence that hybridizes to a nucleotide sequence within any one of nucleotide sequences recited in **TABLE 10** or a reverse complement thereof. A plasmid containing the donor nucleic acid, which encodes any one of amino acid sequences recited in **TABLE 11**, is also provided.

[703] The effector protein system and the plasmid as described above are transfected into human cells via lipofection with a total of 150 ng of each plasmid (300 ng total). Cells are incubated for 48 hours to 72 hours before being harvested for DNA, PCR amplified and sequenced via NGS. The sequencing data are then analyzed using CRISPRESSO to detect/quantify %indel and gene insertion.

Example 5: AAV vectors for gene editing

[704] An AAV vector is constructed to contain a transgene between its ITRs, the transgene providing or encoding, in a 5' to 3' direction, a first promoter, a first guide nucleic acid, a second promoter, an effector protein, a WPRE enhancer, a hGH poly A signal sequence are packaged into the AAV vector. The combination of guide nucleic acid and the effector protein are used for cleaving a target nucleic acid at a specific location. Additionally, the transgene may comprise a donor nucleic acid encoding any one of amino acid sequences recited in **TABLE 11**. Optionally, the AAV vector may have a third promoter and a second guide nucleic acid after the poly A signal sequence in a 5' to 3' direction. The second guide nucleic acid can be same as the first guide nucleic acid where higher efficiency of cleaving is required. The second guide nucleic acid can be different from the first guide nucleic acid where

cleaving is required at two different locations within the target nucleic acid. The effector protein has an amino acid sequence that is at least 90% identical to any one of the sequences recited **TABLE 1A** and **TABLE 1B**. The guide nucleic acid comprises a spacer sequence that hybridizes to a nucleotide sequence within any one of the nucleotide sequences recited in **TABLE 10** or a reverse complement thereof. The effector protein can be expressed either ubiquitously or in a specific tissue based on the promotor the AAV vector is engineered to have. In particular, the second promotor can be ApoE, TBG, MND, CAG, Ck8e, Spc5-12, or Desmin. The first and third promoter can be independently selected from CMV, EF1a, SV40, PGK1, Ubc, human beta actin, CMV, 7SK, EF1a, RPBSA, hPGK, SV40, PGK1, Ubc, human beta actin, CAG, TRE, UAS, Ac5, polyhedron, CaMKIIa, GAL1-10, H1, TEF1, GDS, ADH1, CaMV35S, Ubi, H1, U6, MNDU3, and HSV TK.

Example 6: Targeting intron 1 of human ALB for effector protein CasPhi.12 L26R in primary human hepatocytes

[705] CasPhi.12 system was tested for its indel activity within intron 1 of human albumin gene in an exemplary human primary hepatocytes. Briefly, CasPhi.12 system including the effector protein CasPhi.12 L26R **(SEQ ID NO: 228)** and one of the gRNAs described in **TABLE 16** were used for cleaving within intron 1 of human albumin gene **(SEQ ID NO: 226)** in human primary human hepatocytes. Cas9 system was used as a positive control.

TABLE 16. Exemplary Compositions of Effector Protein and Guide Nucleic Acids

Comp NO:	Effector Protein SEQ ID NO:	gRNA ID NO:	PAM	gRNA Seq ID NO:
48	228	R10006	TTT	652
74	228	R10032	TTC	678
77	228	R10035	TTG	681
81	228	R10039	TTA	685
87	228	R10045	TTT	691
88	228	R10046	TTT	692
93	228	R10051	TTA	697
100	228	R10058	TTC	704
101	228	R10059	TTA	705
109	228	R10067	TTA	713
110	228	R10068	TTT	714
126	228	R10084	TTT	730
127	228	R10085	TTT	731
133	228	R10091	TTA	737
134	228	R10092	TTG	738
136	228	R10094	TTG	740
147	228	R10105	TTT	751
173	228	R10131	TTG	777

Comp NO:	Effector Protein SEQ ID NO:	gRNA ID NO:	PAM	gRNA Seq ID NO:
181	228	R10139	TTT	785
205	228	R10163	TTT	809
208	228	R10166	TTC	812
209	228	R10167	TTT	813

[706] For determining indel activity, 100,000 human primary human hepatocytes were seeded in 24-well collagen I coated plates. After 24 hours, cells were transfected with 2 µg total RNA (1:1 mRNA to gRNA ratio) using Lipofectamine MessengerMax. 72 hours post-transfection, DNA was extracted. DNA editing was quantified by amplicon sequencing.

[707] **FIG. 5** shows % indel activity of CasPhi.12 system in combination with gRNA for cleaving within intron 1 of human Albumin gene, which demonstrates high percentage of total hits as well as high editing efficiency. An analysis of **FIG. 5** indicates that CasPhi.12 system can be used to insert a nucleotide sequence encoding a donor nucleic acid within intron 1 of human Albumin gene. Further analysis of **FIG. 5** indicates that the vast majority of the selected candidate guides were active in human primary human hepatocytes with certain guides reaching levels comparable to Cas9 system.

Example 7: Targeting intron 1 of human *SERPINA1* gene by effector protein CasPhi.12 and guide nucleic acid in primary human hepatocytes

[708] CasPhi.12 system is tested for its indel activity within intron 1 of human *SERPINA1* gene in human primary hepatocytes. Briefly, CasPhi.12 system comprises the effector protein comprising any one of the amino acid sequences of **SEQ ID NO: 6, 228-230** and a gRNA, wherein the gRNA is a crRNA. The crRNA comprises a repeat sequence of **SEQ ID NO: 252** and any one of the spacer sequences of **SEQ ID NO: 911-1612**. An RNP complex of the effector protein and the crRNA recognizes any one of PAM sequences of **SEQ ID NO: 234-237** as identified in **TABLE 17**. A composition comprising the effector protein and the crRNA comprising the repeat sequence and the spacer sequence as described in **TABLE 17** is used for cleaving intron 1 of human *SERPINA1* gene **(SEQ ID NO: 1795)** in human primary human hepatocytes.

TABLE 17. Exemplary Compositions of Effector Protein and crRNA

Effector Protein SEQ ID NO:	PAM Seq ID NO:	Repeat Seq ID NO:	Spacer Seq ID No
6, 228-230	234	252	911-1074
6, 228-230	235	252	1075-1205
6, 228-230	236	252	1206-1432
6, 228-230	237	252	1433-1612

[709] For determining indel activity, 100,000 human primary human hepatocytes are seeded in 24-well collagen I coated plates. After 24 hours, cells were transfected with 2 μg total RNA (1:1 mRNA to gRNA ratio) using Lipofectamine MessengerMax. 72 hours post-transfection, DNA is extracted. DNA editing is quantified by amplicon sequencing.

Example 8: gRNAs targeting intron 1 of human *SERPINA1* for effector protein

[710] CasM.265466 system targeting intron 1 of human *SERPINA1* gene (**SEQ ID NO: 240**) was tested for the ability to produce indels in Heb3B cells. The CasM.265466 system comprised CasM.265466 effector protein (**SEQ ID NO: 231**) and a guide RNA. The guide RNA comprised a handle sequence of **SEQ ID NO: 264**, and one of spacer sequences of **SEQ ID NOS: 911-968** and **1613-1762**.

[711] Briefly, 20,000 cells were seeded in a 96-well plate and transfected using lipofectamine MessengerMax. Hep3B cells were added 200 ng of MessengerMax (1:1 mRNA/gRNA ratio) 24 hours post seeding. The guide RNAs for the ability to produce indels were tested in duplicate. Three SpyCas9 systems were used as positive controls. The cells were harvested, and the media was collected from 48 hours post-transfection. The cells were subject to DNA extraction, PCR amplified and sequenced via NGS sequencing analysis.

[712] **TABLE 18** summarizes average indel data for each of the composition comprising the CasM.265466 effector protein and guide RNA tested.

TABLE 18. Indel Formation using CasM.265466 mRNA paired with various sgRNA in Hep3B cells

Comp. No.	Spacer SEQ ID NO:	Avg. Indel	Comp. No.	Spacer SEQ ID NO:	Avg. Indel	Comp. No.	Spacer SEQ ID NO:	Avg. Indel
R12996	1713	+	R13066	1729	+++	R13135	966	++
R12997	1689	+++	R13067	1642	+++	R13136	919	+
R12998	1613	+	R13068	1698	+	R13137	926	+++
R12999	1614	+	R13069	1643	+	R13138	938	+
R13000	958	+++	R13070	1699	+++	R13139	1662	+++
R13001	1714	+	R13071	1730	++	R13140	947	+
R13002	1615	+	R13072	1700	+	R13141	1705	+++
R13003	1616	++	R13073	1644		R13142	1663	+++
R13004	1690	+	R13074	1731	+	R13143	913	+++
R13005	960	+++	R13075	1645	+++	R13144	1706	+
R13006	957		R13076	1732	++	R13145	1707	+++
R13007	1617	+	R13077	918	+	R13146	914	+++

Comp. No.	Spacer SEQ ID NO:	Avg. Indel
R13008	1691	+
R13009	932	++
R13010	1618	+
R13011	1715	+
R13012	1716	+
R13013	1692	+++
R13014	1619	+++
R13015	961	+
R13016	1717	+
R13017	1620	+++
R13018	1621	+++
R13019	945	+
R13020	1622	+
R13021	1718	+
R13022	1719	+++
R13023	911	+++
R13024	1693	
R13025	1623	+
R13026	1624	+++
R13027	1625	+++
R13028	1720	+
R13029	933	+
R13030	915	
R13031	1626	++
R13032	1627	+
R13033	1721	+++
R13034	1694	+++
R13035	1722	+++
R13036	937	+
R13037	1628	++
R13038	1629	+
R13039	1630	+++
R13040	1631	+
R13041	1632	+++
R13042	1723	+
R13043	1724	+++
R13044	1633	+
R13045	1725	+++
R13046	1726	+++
R13047	1634	+
R13048	917	+++
R13049	1635	+++
R13050	948	+
R13051	1636	++
R13052	964	+++
R13053	1637	+
R13054	967	++
R13055	921	+++
R13056	1727	+++

Comp. No.	Spacer SEQ ID NO:	Avg. Indel
R13078	931	+
R13079	942	+++
R13080	922	+++
R13081	963	+++
R13082	1646	+++
R13083	1701	+++
R13084	1733	+++
R13085	965	
R13086	912	+++
R13087	1734	+
R13088	950	+++
R13089	1647	+
R13090	1648	+++
R13091	1649	++
R13092	941	
R13093	929	+
R13094	1650	+++
R13095	1651	
R13096	1735	
R13097	936	+++
R13098	928	+
R13099	1736	
R13100	1737	+
R13101	1652	+
R13102	959	
R13103	1702	+
R13104	1653	
R13105	1654	+++
R13106	924	
R13107	1655	++
R13108	1738	+
R13109	1703	+++
R13110	956	++
R13111	955	+++
R13112	1704	+++
R13113	935	+
R13114	1656	+++
R13115	1739	+++
R13116	952	
R13117	940	+++
R13118	1740	+++
R13119	1657	++
R13120	1741	+++
R13121	1658	+
R13122	1742	
R13123	1743	++
R13124	951	+
R13125	1744	+
R13126	927	

Comp. No.	Spacer SEQ ID NO:	Avg. Indel
R13147	925	
R13148	1748	
R13149	1664	++
R13150	1665	+
R13151	954	+
R13152	962	
R13153	1666	+++
R13154	1708	+++
R13155	1749	+++
R13156	1750	+++
R13157	1667	+++
R13158	1751	+
R13159	930	+++
R13160	1709	++
R13161	1752	+++
R13162	1753	
R13163	1754	+++
R13164	1668	+
R13165	1669	++
R13166	1670	+++
R13167	1710	++
R13168	1711	
R13169	1671	+
R13170	1755	++
R13171	1756	+++
R13172	1672	++
R13173	1757	+
R13174	1712	+
R13175	920	
R13176	949	+
R13177	1673	+++
R13178	968	+++
R13179	1758	+++
R13180	1674	+
R13181	1675	+++
R13182	1676	+++
R13183	944	+
R13184	1677	+
R13185	1678	
R13186	1679	+
R13187	1680	+++
R13188	1681	+++
R13189	1759	+++
R13190	1760	+++
R13191	1682	+++
R13192	1683	+
R13193	1684	+++
R13194	916	++
R13195	943	+

Comp. No.	Spacer SEQ ID NO:	Avg. Indel
R13057	1638	+++
R13058	946	++
R13059	1695	+++
R13060	1696	+
R13061	1639	+++
R13062	1697	+
R13063	1640	+
R13064	1641	++
R13065	1728	++

Comp. No.	Spacer SEQ ID NO:	Avg. Indel
R13127	1745	
R13128	1746	
R13129	923	++
R13130	1659	+
R13131	934	
R13132	1747	++
R13133	1660	+++
R13134	1661	+

Comp. No.	Spacer SEQ ID NO:	Avg. Indel
R13196	1685	+
R13197	1686	++
R13198	1687	+
R13199	953	++
R13200	1761	+
R13201	1688	++
R13202	1762	
R13203	939	+++

NOTE: Magnitude of data: "+++ " represents a value >10, " ++ " represents a value between ≤10 and ≥5, "+ " represents a value <5.

[713] An analysis of indel results summarized in **TABLE 18** indicates that intron 1 of human *SERPINA1* gene can be modified using Cas265466 system.

Example 9: Targeting intron 1 of human *SERPINA1* gene by effector protein Cas14a.1 and guide nucleic acid in primary human hepatocytes

[714] Cas14a.1 system is tested for its indel activity within intron 1 of human *SERPINA1* gene in human primary hepatocytes. Briefly, Cas14a.1 system comprises the effector protein comprising an amino acid sequence of **SEQ ID NO: 13** and a gRNA, wherein the gRNA is a sgRNA. The sgRNA comprises any one of handle sequences of **SEQ ID NO: 259-261** and any one of the spacer sequences of **SEQ ID NO: 911-968**, and **1100-1143**. An RNP complex of the effector protein and the sgRNA recognizes a PAM sequence of **SEQ ID NO: 238 or 247** as identified in **TABLE 19**. A composition comprising the effector protein, and the sgRNA comprising the handle sequence and the spacer sequence as described in **TABLE 19** is used for cleaving intron 1 of human *SERPINA1* gene **(SEQ ID NO: 1795)** in human primary human hepatocytes.

TABLE 19. Exemplary Compositions of Effector Protein and sgRNA

Effector Protein SEQ ID NO:	PAM Seq ID NO:	Handle Seq ID NO:	Spacer Seq ID No
13	247	259-261	911-968
13	238	259-261	1100-1143

[715] For determining indel activity, 100,000 human primary human hepatocytes are seeded in 24-well collagen I coated plates. After 24 hours, cells are transfected with 2 µg total RNA (1:1 mRNA to gRNA ratio) using Lipofectamine MessengerMax. 72 hours post-transfection, DNA is extracted. DNA editing is quantified by amplicon sequencing.

Example 10: Targeting intron 1 of human *SERPINA1* gene by effector protein CasPhi.32 and guide nucleic acid in primary human hepatocytes

[716] CasPhi.32 system is tested for its indel activity within intron 1 of human *SERPINA1* gene in human primary hepatocytes. Briefly, CasPhi.32 system comprises the effector protein comprising an amino acid sequence of **SEQ ID NO: 47** and a gRNA, wherein the gRNA is a crRNA. The crRNA comprises a repeat sequence of **SEQ ID NO: 254** and any one of the spacer sequences of **SEQ ID NO: 969-999**, **1075-1099**, **1206-1251**, and **1583-1612**. An RNP complex of the effector protein and the crRNA recognizes any one of PAM sequences of **SEQ ID NO: 241-244** as identified in **TABLE 20**. A composition comprising the effector protein, and the crRNA comprising the repeat sequence and the spacer sequence as described in **TABLE 20** is used for cleaving intron 1 of human *SERPINA1* gene **(SEQ ID NO: 1795)** in human primary human hepatocytes.

TABLE 20. Exemplary Compositions of Effector Protein and crRNA

Effector Protein SEQ ID NO:	PAM Seq ID NO:	Repeat Seq ID NO:	Spacer Seq ID No
47	241	254	969-999
47	244	254	1075-1099
47	243	254	1206-1251
47	242	254	1583-1612

[717] For determining indel activity, 100,000 human primary human hepatocytes are seeded in 24-well collagen I coated plates. After 24 hours, cells are transfected with 2 μg total RNA (1:1 mRNA to gRNA ratio) using Lipofectamine MessengerMax. 72 hours post-transfection, DNA is extracted. DNA editing is quantified by amplicon sequencing.

Example 11: Targeting intron 1 of human *SERPINA1* gene by effector protein CasM.19952 and guide nucleic acid in primary human hepatocytes

[718] CasM.19952 system is tested for its indel activity within intron 1 of human *SERPINA1* gene in human primary hepatocytes. Briefly, CasM.19952 system comprises the effector protein comprising any one of the amino acid sequences of **SEQ ID NO: 181-184** and a gRNA, wherein the gRNA is a sgRNA. The sgRNA comprises any one of handle sequences of **SEQ ID NO: 262-264** and any one of the spacer sequences of **SEQ ID NO: 1763-1788**. An RNP complex of the effector protein and the sgRNA recognizes any one of PAM sequences of **SEQ ID NO: 1791-1794** as identified in **TABLE 21**. A composition comprising the effector protein, and the sgRNA comprising the handle sequence and the spacer sequence as described in **TABLE 21** is used for cleaving intron 1 of human *SERPINA1* gene **(SEQ ID NO: 1795)** in human primary human hepatocytes.

TABLE 21. Exemplary Compositions of Effector Protein and crRNA

Effector Protein SEQ ID NO:	PAM Seq ID NO:	Handle Seq ID NO:	Spacer Seq ID No
181-184	1791	262-264	1763-1766
181-184	1792	262-264	1767-1781
181-184	1793	262-264	1782-1786
181-184	1794	262-264	1787-1788

[719] For determining indel activity, 100,000 human primary human hepatocytes are seeded in 24-well collagen I coated plates. After 24 hours, cells are transfected with 2 µg total RNA (1:1 mRNA to gRNA ratio) using Lipofectamine MessengerMax. 72 hours post-transfection, DNA is extracted. DNA editing is quantified by amplicon sequencing.

Example 12: CasPhi.12 mediated gene editing in primary human hepatocytes derived from three donors

[720] 100,000 primary human hepatocytes from three donors (designated ERR, OGF, and UNR) were transfected with 400, 100, 25, or 5 ng mRNA encoding effector protein of **SEQ ID NO: 6** and gRNA at 1:1 ratio by MessengerMax while rocking in low adhesion 96-well plates for 2 hours. The cells were then transferred to BioCoat collagen I 96-well plates for culture. After 48 hours, viability was measured by MTT assay and editing assessed by NGS. The results are summarized in **TABLE 22, TABLE 23** and **TABLE 24**.

TABLE 22. Result of % Indel generated in Primary Human Hepatocytes of ERR donor by mRNA encoding Effector Protein and gRNA

Effector Protein	Guide RNA	ERR							
		400 ng		100 ng		25 ng		5 ng	
Cas9	G009874	ND	+++	++	+++	+	++	+	+
SEQ ID NO: 6	**SEQ ID NO: 625**	++	++	+	+	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 652**	+++	+++	++	++	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 653**	++	++	+	+	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 685**	+++	+++	++	+	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 690**	+	+	+	+	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 705**	+++	++	++	++	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 706**	++	+++	++	++	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 713**	+++	+++	++	++	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 714**	++	+++	++	++	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 727**	++	+++	++	++	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 774**	++	++	++	++	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 777**	+++	+++	++	+	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 785**	+	+	+	+	+	+	+	+

SEQ ID NO: 6	**SEQ ID NO: 790**	+++	+++	++	+	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 808**	ND	+	+	+	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 809**	+	+	ND	+	+	+	+	+

"+" indicates <5% indel; "++" indicates ≥5% to <30% indel; "+++" indicates ≥30% indel; ND = Not detected.

TABLE 23. Result of % Indel generated in Primary Human Hepatocytes of OGF donor by mRNA encoding Effector Protein and gRNA

Effector Protein	**Guide RNA**	**OGF**							
		400 ng		**100 ng**		**25 ng**		**5 ng**	
Cas9	G009874	+++	+++	++	++	++	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 625**	++	+	+	+	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 652**	+++	++	+	++	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 653**	++	+	+	+	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 685**	++	++	++	+	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 690**	+	+	+	+	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 705**	+	++	++	+	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 706**	+	++	+	+	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 713**	+	++	+	++	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 714**	++	++	+	+	+	+	ND	+
SEQ ID NO: 6	**SEQ ID NO: 727**	+	++	+	+	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 774**	+	++	+	+	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 777**	+++	++	++	++	++	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 785**	++	+	+	+	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 790**	++	++	+	+	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 808**	+	+	+	+	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 809**	+	+	+	+	+	+	+	+

"+" indicates <5% indel; "++" indicates ≥5% to <30% indel; "+++" indicates ≥30% indel; ND = Not detected.

TABLE 24. Result of % Indel generated in Primary Human Hepatocytes of UNR donor by mRNA encoding Effector Protein and gRNA

Effector Protein	**Guide RNA**	**UNR**							
		400 ng		**100 ng**		**25 ng**		**5 ng**	
Cas9	G009874	+++	+++	++	++	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 625**	+	+	+	+	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 652**	++	++	++	++	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 653**	+	+	+	+	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 685**	++	++	+	+	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 690**	+	+	+	+	+	+	+	+

SEQ ID NO: 6	**SEQ ID NO: 705**	++	++	+	+	+	+	+	ND
SEQ ID NO: 6	**SEQ ID NO: 706**	++	++	+	+	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 713**	++	++	+	+	+	+	+	ND
SEQ ID NO: 6	**SEQ ID NO: 714**	++	++	+	+	+	+	+	ND
SEQ ID NO: 6	**SEQ ID NO: 727**	++	++	+	+	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 774**	++	++	+	+	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 777**	++	+++	++	++	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 785**	+	+	+	+	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 790**	++	++	+	+	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 808**	+	+	+	+	+	+	+	+
SEQ ID NO: 6	**SEQ ID NO: 809**	+	+	+	+	+	+	+	+

"+" indicates <5% indel; "++" indicates ≥5% to <30% indel; "+++" indicates ≥30% indel; ND = Not detected.

Example 13: CasPhi.12 mediated gene editing in primary human hepatocytes derived from a donor

[721] 500,000 primary human hepatocytes from the donor ERR were transfected with 2 µg of mRNA encoding effector protein of **SEQ ID NO: 6** and gRNA at 1:1 ratio by MessengerMax while rocking in low adhesion 24-well plates for 2 hours. The cells were then transferred to Collagen I coated 24-well plate for culture. After 48 hours, viability was measured by MTT assay and editing assessed by NGS. The results are summarized in **TABLE 25**.

TABLE 25. Result of % Indel generated in Primary Human Hepatocytes by mRNA encoding Effector Protein and gRNA

Effector Protein	**Guide RNA**	**ERR**	
		% indel	
Cas9	G009874	+++	+++
SEQ ID NO: 6	**SEQ ID NO: 625**	++	++
SEQ ID NO: 6	**SEQ ID NO: 652**	+++	+++
SEQ ID NO: 6	**SEQ ID NO: 653**	++	++
SEQ ID NO: 6	**SEQ ID NO: 685**	++	++
SEQ ID NO: 6	**SEQ ID NO: 690**	+	+
SEQ ID NO: 6	**SEQ ID NO: 705**	+++	+++
SEQ ID NO: 6	**SEQ ID NO: 706**	+++	+++
SEQ ID NO: 6	**SEQ ID NO: 713**	+++	+++
SEQ ID NO: 6	**SEQ ID NO: 714**	+++	+++

SEQ ID NO: 6	**SEQ ID NO: 727**	++	++
SEQ ID NO: 6	**SEQ ID NO: 774**	++	++
SEQ ID NO: 6	**SEQ ID NO: 777**	++	+
SEQ ID NO: 6	**SEQ ID NO: 785**	+++	++
SEQ ID NO: 6	**SEQ ID NO: 790**	+++	++
SEQ ID NO: 6	**SEQ ID NO: 808**	+	+
SEQ ID NO: 6	**SEQ ID NO: 809**	+	+

"+" indicates <5% indel; "++" indicates ≥5% to <60% indel; "+++" indicates ≥60% indel; ND = Not detected.

Example 14: CasM.265466 mediated gene editing in primary human hepatocytes derived from three donors

[722] 100,000 primary human hepatocytes from three donors (designated ERR, OGF, and UNR) were transfected with 400, 100, 25, or 5 ng mRNA encoding effector protein of **SEQ ID NO: 231** and gRNA at 1:1 ratio by MessengerMax while rocking in low adhesion 96-well plates for 2 hours. The cells were then transferred to BioCoat collagen I 96-well plates for culture. After 48 hours, viability was measured by MTT assay and editing assessed by NGS. The results are summarized in **TABLE 26, TABLE 27** and **TABLE 28**.

TABLE 26. Result of % indel generated in primary human hepatocytes of ERR donor by mRNA encoding effector protein and gRNA

Effector Protein	**Guide RNA**	**ERR**							
		400 ng		**100 ng**		**25 ng**		**5 ng**	
Cas9	G009874	ND	+++	++	+++	+	++	+	+
SEQ ID NO: 231	**SEQ ID NO: 856**	+++	+++	++	++	++	++	+	+
SEQ ID NO: 231	**SEQ ID NO: 858**	++	++	+	+	+	+	+	+
SEQ ID NO: 231	**SEQ ID NO: 862**	++	++	+	+	+	+	+	+
SEQ ID NO: 231	**SEQ ID NO: 864**	+++	+++	++	++	+	+	+	+
SEQ ID NO: 231	**SEQ ID NO: 866**	+++	+++	++	++	++	+	+	+
SEQ ID NO: 231	**SEQ ID NO: 867**	++	++	+	+	+	+	+	+
SEQ ID NO: 231	**SEQ ID NO: 868**	++	++	++	++	++	+	+	+

"+" indicates <5% indel; "++" indicates ≥5% to <30% indel; "+++" indicates ≥30% indel; ND = Not detected.

TABLE 27. Result of % indel generated in primary human hepatocytes of OGF donor by mRNA encoding effector protein and gRNA

Effector Protein	Guide RNA	OGF							
		400 ng		100 ng		25 ng		5 ng	
Cas9	G009874	+++	++	++	++	+	+	+	+
SEQ ID NO: 231	**SEQ ID NO: 856**	+	+	+	+	+	+	+	+
SEQ ID NO: 231	**SEQ ID NO: 858**	+	+	+	+	+	+	+	+
SEQ ID NO: 231	**SEQ ID NO: 862**	++	++	+	+	+	+	+	+
SEQ ID NO: 231	**SEQ ID NO: 864**	++	++	+	++	+	+	+	+
SEQ ID NO: 231	**SEQ ID NO: 866**	+	+	+	+	+	+	+	+
SEQ ID NO: 231	**SEQ ID NO: 867**	++	++	+	+	+	+	+	+
SEQ ID NO: 231	**SEQ ID NO: 868**	+++	++	++	++	+	+	+	+

"+" indicates <5% indel; "++" indicates ≥5% to <30% indel; "+++" indicates ≥30% indel; ND = Not detected.

TABLE 28. Result of % indel generated in primary human hepatocytes of UNR donor by mRNA encoding effector protein and gRNA

Effector Protein	Guide RNA	UNR							
		400 ng		100 ng		25 ng		5 ng	
Cas9	G009874	+++	+++	++	++	+	+	+	+
SEQ ID NO: 231	**SEQ ID NO: 856**	+++	+++	++	++	+	+	+	+
SEQ ID NO: 231	**SEQ ID NO: 858**	++	++	++	+	+	+	+	+
SEQ ID NO: 231	**SEQ ID NO: 862**	+	+	+	+	+	+	ND	+
SEQ ID NO: 231	**SEQ ID NO: 864**	+	+	+	+	+	+	+	+
SEQ ID NO: 231	**SEQ ID NO: 866**	++	++	+	+	+	+	+	+
SEQ ID NO: 231	**SEQ ID NO: 867**	++	++	+	+	+	+	+	+
SEQ ID NO: 231	**SEQ ID NO: 868**	+	+	+	+	+	+	+	+

"+" indicates <5% indel; "++" indicates ≥5% to <30% indel; "+++" indicates ≥30% indel; ND = Not detected.

Example 15: CasM.265466 mediated gene editing in primary human hepatocytes derived from a donor

[723] 500,000 primary human hepatocytes from the donor ERR were transfected with 2 μg of mRNA encoding effector protein of **SEQ ID NO: 231** and gRNA at 1:1 ratio by MessengerMax while rocking in low adhesion 24-well plates for 2 hours. The cells were then transferred to Collagen I coated 24-well plate for culture. After 48 hours, viability was measured by MTT assay and editing assessed by NGS. The results are summarized in **TABLE 29**.

TABLE 29. Result of % indel generated in primary human hepatocytes by mRNA encoding effector protein and gRNA

Effector Protein	Guide RNA	ERR	
		% indel	
Cas9	G009874	+++	+++

SEQ ID NO: 231	**SEQ ID NO: 856**	+++	+++
SEQ ID NO: 231	**SEQ ID NO: 858**	++	++
SEQ ID NO: 231	**SEQ ID NO: 862**	++	++
SEQ ID NO: 231	**SEQ ID NO: 864**	+++	++
SEQ ID NO: 231	**SEQ ID NO: 866**	+++	+++
SEQ ID NO: 231	**SEQ ID NO: 867**	++	++
SEQ ID NO: 231	**SEQ ID NO: 868**	++	+

"+" indicates <10% indel; "++" indicates ≥10% to <50% indel; "+++" indicates ≥50% indel; ND = Not detected.

Example 16: PAM Screening for D2S effector protein CasM.265466

[724] D2S effector protein CasM.265466 and guide RNA combinations represented in **TABLE 30** were screened by *in vitro* enrichment (IVE) for PAM recognition. **TABLE 30** show the components of each effector protein-guide RNA complex assayed for PAM recognition. The amino acid sequence of CasM.265466 is recited in **SEQ ID NO: 231**. The nucleotide sequences of the guide components are shown in **TABLE 30** herein. For example, as shown in **TABLE 30**, an effector protein comprising an amino acid sequence of **SEQ ID NO: 231** complexed with a guide comprising a crRNA of **SEQ ID NO: 1825** and a tracrRNA sequence of **SEQ ID NO: 1826** was screened for PAM recognition.

TABLE 30: Exemplary Compositions used for PAM Screening

No.	Composition	Sequence
1	Effector Protein	CasM.265466 **(SEQ ID NO: 231)**
	crRNA	GUUUGAGAACCUUAUGAAAUUACAAGGAUGCCAAACUAUUAA AUACUCGUAUUGCU **(SEQ ID NO: 1825)**
	tracrRNA sequence	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCA CAAGAAUCCU **(SEQ ID NO: 1826)**
2	Effector Protein	CasM.265466 **(SEQ ID NO: 231)**
	crRNA	GUUUGAGAACCUUAUGAAAUUACAAGGAUGCCAAACUAUUAA AUACUCGUAUUGCU **(SEQ ID NO: 1825)**
	tracrRNA sequence	UAUAUUUGAUAAAAAUAUACAGCUUAUUUGGAAGCUGAAAUG UGAGGUUUAUAACACUCACAAGAAUCC **(SEQ ID NO: 1827)**
3	Effector Protein	CasM.265466 **(SEQ ID NO: 231)**
	sgRNA	ACAGCUUAUUUGGAAGCUGAAAUGUGAGGUUUAUAACACUCA CAAGAAUCCUGAAAAAGGAUGCCAAACUAUUAAAUACUCGUA UUGCU **(SEQ ID NO: 1828)**

[725] Briefly, effector proteins were complexed with corresponding guide RNAs for 15 minutes at 37 °C. The complexes were added to an IVE reaction mix. PAM screening reactions used 10 µl of RNP in 100 µl reactions with 1,000 ng of a 5' PAM library in 1x Cutsmart buffer and were carried out for 15 minutes at 25 °C, 45 minutes at 37 °C and 15 minutes at 45 °C. Reactions were terminated with 1 µl of proteinase K and 5 µl of 500 mM EDTA for 30 minutes at 37 °C. Next generation sequencing was

performed on cut sequences to identify enriched PAM sequence for CasM.265466. *Cis* cleavage by each complex was confirmed by gel electrophoresis. The most enriched PAM was represented by the sequence 5'-TNTR-3' (**SEQ ID NO: 1829**), wherein N is any nucleotide and R is adenine or guanine.

Example 17: *In vitro* enrichment in mammalian cells

[726] In this experiment CasM.265466 was expressed in HEK293T cells, and cell lysate was tested for nucleic acid cleaving activity. Purified CasM.265466 from HEK293T cells was also tested for nucleic acid cleaving activity. In both cases, *cis* cleavage activity was detected by the presence of bands. The PAM requirements were determined to be TNTR (**SEQ ID NO: 1829**) by NGS after *in vitro* enrichment of DNA fragments.

[727] *In vitro* enrichment involved the amplification of DNA fragments excised by potential CRISPR-Cas candidates. The method began with a *cis* cleavage assay, which was then followed by dA end repair, ligation, and multiple rounds of PCR. Magnetic bead purification was also performed after interference, ligation, and both rounds of PCR. The final purified PCR product was then sequenced on a MiSeq instrument. Details of these steps are provided as follows.

HEK293 Lipofection and Lysis

[728] Opti-MEM media was warmed to 37 °C and transfection reagent was equilibrated at room temperature. Final transfection ratio was prepared at pDNA:Lipid – 300 ng pDNA : 0.6 µl tx reagent per transfection. A first solution was prepared by diluting pDNA with Opti-MEM – 360 ng pDNA diluted with media to final volume of 12 µL. A second solution was prepared by diluting the transfection reagent with Opti-MEM - 0.72 µl tx reagent diluted with media to final volume of 12 µL. 12 µL of the first solution and 12 µL of the second solution were mixed and incubated at room temperature for 15 minutes, and then a 20 µL aliquot of the mixture was dispensed over the cells followed by incubation at 37 °C for approximately 72 hours before harvesting.

Interference Assay

[729] Purified CRISPR effector protein CasM.265466 (50 µM) was added to a reaction containing 10 µl 10x Cutsmart buffer, and a plasmid (1000 ng per reaction). Additionally, prepared in parallel was a solution containing 3 µL of EcoRI and 7 µL dH_2O as a positive control. Dilutions and volumes of the prepared reactions for 3' PAM and 5' PAM are shown in **TABLE 31**. The reaction was incubated at 37 °C for 30 minutes, 5 µL EDTA + 1 µL Proteinase K solution were then added, and the reaction was further incubated at 37 °C for 30 minutes. NGS was subsequently performed, and the required PAM determined was 5'-TNTR-3' (**SEQ ID NO: 1829**).

TABLE 31: Interference Assay Reaction

	3'PAM		5'PAM	
	1x Volume (µL)	50	1X Volume (uL)	20
10x CutSmart Buffer	10	500	10	200
Plasmid (1000ng/rxn)	0.9496676163	47.4833808167	3.0303030303	60.6060606061
dH2O		3852.5166191833		1499.3939393939
Protein (50 µM)	12	600	12	240
Total	100	5000	100	2000

Magnetic Bead Purification I

[730] SPRIselect beads for resuspension in solution were prepared. 60 µL of each bead solution was added to each interference assay reaction and incubated for 5 minutes at 25 °C. The reactions were then placed on a magnetic stand. After 1 minute, clear liquid was aspirated from each reaction without disturbing the magnetic beads. To each reaction still containing the magnetic beads, 190 µL of 80% ethanol was added. The ethanol was then removed from each well without disturbing the magnetic beads. The addition and removal of ethanol was repeated with 200 µL of 80% ethanol to each reaction. The magnetic beads in each reaction vessel were then resuspended in 55 µL nuclease free 1x TE buffer or dH_2O. The resuspension solutions were incubated for 1 minute at 25 °C and returned to the magnetic stand. 50 µL of each resuspension solution were then transferred into new reaction plates.

End Repair – dA-Tailing

[731] Reactions containing purified DNA were expose to 7 µL of Ultra II EP buffer and diluted to 180 µL. An additional 3 µL of Ultra II End Prep Enzyme Mix was added to each reaction. The reactions were then mixed thoroughly and then placed in a thermocycler according to the timeline in **TABLE 32**.

TABLE 32: Thermocycler Programming

Steps	Time (minutes)	Temperature (°C)
1	30	20
2	30	65
3	∞	4

Adapter Ligation

[732] To the end prepared reactions described above, the following components described in **TABLE 33** were added at 0 °C. An adapter sequence of ILM8-UDI-UMI was used and the reactions were mixed thoroughly and incubated at 20 °C for 15 minutes in a thermocycler with the heated lid removed from the apparatus.

TABLE 33: ADAPTER Ligation Reaction Components

	Volume (µL)
EndRepair Reaction	60
NEBNext Ultra II Ligation Master Mix	30
NEBNext Ligation Enhancer	1
ILM8-UDI-UMI adaptor (1.5 µM)	2.5
dH2O	6.5
Total	100

Magnetic Bead Purification II

[733] SPRIselect Beads were mixed to resuspend the beads in solution. To each ligation reaction described above, 60 µL of the SPRIselect Bead suspension were added and then 25 µL of nuclease-free water was added.

PCR for Target Enrichment

[734] PCR for target enrichment was conducted by preparing reactions with various IVE primers possessing different overhang sequences as shown in **TABLE 34** and **TABLE 35**.

TABLE 34: PCR for TARGET Enrichment Reaction Components

Reagent	1x Volume (µL)	96
2X Q5 NEBNext	12.5	1200
IVE F pool (100 µM)	0.125	12
P7 Reverse Primer (100 µM)	0.125	12
Water	2.25	216
Post Clean up Ligation Reaction	10	
Total Volume	25	2400

TABLE 35: IVE Primer Sequences

Name	Sequence
IVE longF-A AJD001	ACACTCTTTCCCTACACGACGCTCTTCCGATCtNcctttcgtctcgcgcgtttcgg **(SEQ ID NO: 1830)**
IVE longF-B AJD001	ACACTCTTTCCCTACACGACGCTCTTCCGATCtNNNcctttcgtctcgcgcgtttcg g **(SEQ ID NO: 1831)**
IVE longF-C AJD001	ACACTCTTTCCCTACACGACGCTCTTCCGATCtNNNNNcctttcgtctcgcgcgt ttcgg **(SEQ ID NO: 1832)**
IVE longF-D AJD001	ACACTCTTTCCCTACACGACGCTCTTCCGATCtNNNNNNNcctttcgtctcgc gcgtttcgg **(SEQ ID NO: 1833)**

Example 18: Additional PAM Screening for CasM.265466

[735] Prior *in vitro* screening as described in Example 16 for effector protein CasM.265466 (**SEQ ID NO: 231**) PAM recognition demonstrated that the most enriched PAM sequence for CasM.265466 (**SEQ ID NO: 231**) was a TNTR (**SEQ ID NO: 1829**) PAM sequence, but also indicated that the effector protein may tolerate a more flexible PAM sequences beyond TNTR (**SEQ ID NO: 1829**) without significantly compromising nuclease activity. Effector protein and flexible PAM group combinations as set forth in **TABLE 36** were screened to confirm that chromosomal DNA may be efficiently targeted in mammalian cells (HEK293T) using a more flexible PAM sequence.

[736] Single and double point mutations were made along TNTR (**SEQ ID NO: 1829**).

TABLE 36: PAM sequences

SEQ ID NO:	PAM Group*
1834	NNTN
1835	ANTR
1836	CNTR
1837	GNTR
1838	TNAR
1839	TNCR
1840	TNGR
1841	TNTC
1842	TNTT
1843	VNTY
1844	TNVY

*wherein each N is any nucleotide, each R is A or G, and each V is A, C or G.

[737] At least six spacers that previously showed >3% indel rate were selected for each PAM group identified in **TABLE 36**.

[738] Single guide nucleic acids (sgRNA) comprising a handle sequence of **SEQ ID NO: 264** and a spacer sequence comprising 20 nucleotides was used.

[739] Plasmids encoding CasM.265466 effector protein (**SEQ ID NO: 231**) and plasmids encoding the sgRNAs were delivered via lipofection to HEK293T cells and permitted to grow to allow for indel formation. Cells were lysed and indels were detected by next generation sequencing. Indel percentage was calculated and plotted as shown in **FIG. 6**.

[740] While the top performing complexes were found to produce up to or greater than 30% indel, the data also demonstrates that single and double point mutations at -4 and -1 were the most permissive for allowing nuclease activity. Furthermore, the CasM.265466 effector protein (**SEQ ID NO: 231**) complexed with two different sgRNAs having different spacer sequences generated 20% indel at targeted sequences adjacent to an NNTN (**SEQ ID NO: 1834**) PAM. Therefore, these results further confirm the results of Example 9 and demonstrate that the CasM.265466 effector protein (**SEQ ID NO: 231**) recognizes a flexible NNTN (**SEQ ID NO: 1834**) PAM sequence.

Example 19. Arginine mutation scanning of CasM.265466 to identify charge substitution rules of effector protein activity

[741] CasM.265466 arginine mutants were tested for their ability to produce indels in HEK293T cells. A total of 368 arginine mutants were tested. Briefly, a first plasmid encoding a CasM.265466 arginine mutant and a second plasmid encoding a single guide RNA were delivered by lipofection to HEK293T cells. The sgRNA comprised a handle sequence of **SEQ ID NO: 264** and a spacer sequence of UCUUCGCCCAGAGCAUCCCA **(SEQ ID NO: 1845)**. The sgRNA comprised a spacer sequence that was designed to hybridize to a target sequence adjacent to a PAM of TNTR **(SEQ ID NO: 1829)** (*e.g.*, TTTG **(SEQ ID NO: 247)**). For lipofections, 15 ng of the nuclease mutant and 150 ng of the guide RNA encoding plasmid were delivered to ~30,000 HEK293T cells in 200 µl using TransIT-293 lipofection reagent. Lipofected cells were grown for ~72 hrs at 37 °C to allow for indel formation. Indels were detected by next generation sequencing of PCR amplicons at the targeted loci and indel percentage was calculated as the fraction of sequencing reads containing insertions or deletions relative to an unedited reference sequence. Sequencing libraries with less than 20% of reads aligning to the reference sequence were excluded from the analysis for quality control purposes. Wildtype CasM.265466 was included as positive control and reference for the mutants.

[742] The mean indel percentage for each of the arginine mutant is shown in **FIG. 7.** An analysis of **FIG. 7** indicates that positive charge of arginine may strengthen the interaction between the effector protein and the negatively charged DNA backbone. Top 10 arginine mutants that showed increase in indel potency includes I80R, T84R, K105R, G210R, C202R, A218R, D220R, E225R, C246R, and Q360R.

Example 20. CasM.265466 arginine mutants and their potency for indel generation

[743] The top ten nuclease mutants, each comprising different CasM.265466 arginine mutant, as identified in Example 19 were tested for their ability to produce indels in HEK293T cells over a variety of doses. Briefly, a first plasmid encoding a CasM.265466 mutant and a second plasmid encoding a single guide RNA (sgRNA) were delivered by lipofection to HEK293T cells. The sequence of the sgRNAs comprised a handle sequence of **SEQ ID NO: 264** and a spacer sequence of **SEQ ID NO: 1845**. The sgRNA spacer was designed to hybridize to a target sequence adjacent to a PAM of TNTR **(SEQ ID NO: 1829)** (*e.g.*, TTTG **(SEQ ID NO: 247)**). For lipofections, the CasM.265466 mutant and sgRNA were delivered to ~30,000 HEK293T cells in 200 µl using TransIT-293 lipofection reagent. Each of the ten nuclease mutants were tested at a dose ranging from 1.17 ng to 150 ng. The sgRNA encoding plasmid was used at a concentration of 150 ng. Lipofected cells were grown for ~72 hrs at 37 °C to allow for indel formation. Indels were detected by next generation sequencing of PCR amplicons at the targeted loci and indel percentage was calculated as the fraction of sequencing reads containing insertions or deletions relative to an unedited reference sequence. Sequencing libraries with less than

20% of reads aligning to the reference sequence were excluded from the analysis for quality control purposes. Wildtype CasM.265466 was included as positive control and reference for the mutants.

[744] The mean indel percentage and standard deviation based on three replicates is reported in **FIG. 8**. An analysis of **FIG. 8** indicates that arginine substitution can increase potency of the effector protein in the generation of indels.

Example 21: Editing efficiency and integration rates of CasM.265466 effector protein variants in HEK293T cells

[745] CasM.265466 effector protein (**SEQ ID NO: 231**) and CasM.265466 effector protein variants having either a D220R or a E225R substitution was tested for their editing efficiency and ability to integrate double-stranded oligodeoxynucleotide (dsODN) in HEK293T cells as compared to Cas9, empty plasmids and non-targeting (NT) controls.

[746] Briefly, 1 μg of Cas9 and an equivalent amount of the CasM.265466 effector protein or variants thereof (**0.63** μg) were complexed with guide nucleic acids (**0.37** μg) targeting *MLH1*, and delivered by nucleofection to HEK293T cells (2.3E+05 cells). Likewise, the controls, empty plasmid, Cas9 and NT controls, were complexed with 2, 1, and 0 μg of guide nucleic acids respectively, and also delivered as described. Guide sequence for CasM.265466 WT and D220R variant comprises a handle sequence of **SEQ ID NO: 264** and a spacer sequence comprising a nucleotide sequence of AGUCUCCAGGAAGAAAUUAA (**SEQ ID NO: 1846**). Editing efficiency and integration were assessed by NGS.

[747] The results confirm the findings from Example 19, namely that the variant effector proteins continue to have an increased editing efficiency as compared to the CasM.265466 effector protein, and had an editing efficiency comparable to Cas9 (**FIG. 9**). The editing efficiency of the E225R variant was comparable to the editing efficiency of the D220R variant. The D220R and E225R variants were also capable of integrating the dsODN.

Example 22: Engineered variants of CasM.265466

[748] A potency assay was performed to evaluate the activity of CasM.265466 protein and engineered variants thereof. The variants were identical to CasM.265466 protein with the exception of the following amino acid substitutions: D220R and A306K. HEK293T cells were transfected with plasmids encoding these proteins and a single guide nucleic acid targeting *MLH1*. Wildtype CasM.265466 and Cas9 were included as controls. The nucleotide sequence of the sgRNA used with CasM.265466 and variants thereof comprised a handle sequence of **SEQ ID NO: 264** and a spacer sequence of **SEQ ID NO: 1846**.

[749] The percentage of target nucleic acids exhibiting nucleotide insertion(s)/deletion(s) (% indel), indicative of nuclease activity, was assessed with next generation sequencing (NGS). As shown in **FIG.**

10, D220R and A306K were able to generate higher indel percentage relative to the corresponding wildtype CasM.265466 protein.

Example 23: Rational Engineering of CasM. 265466 variants

[750] CasM.265466 protein and engineered variants thereof were tested for their ability to produce indels in a mammalian cell line (*e.g.*, HEK293T cells). Briefly, a plasmid encoding the effector proteins and a guide RNA were delivered by lipofection to the mammalian cells. This was performed with guide RNA comprising a handle sequence of **SEQ ID NO: 264** and a spacer sequence of **SEQ ID NO: 1846**, wherein the guide RNA was targeting loci adjacent to a PAM of TNTR. Indels in the loci were detected by next generation sequencing of PCR amplicons at the targeted loci and % indel was calculated as the fraction of sequencing reads containing insertions or deletions relative to an unedited reference sequence. **FIG. 11** shows the activity of the engineered variants relative to that of wildtype as fold change. **TABLE 37** lists CasM.265466 variants that demonstrated increased indel production relative to wildtype CasM.265466 (starting with highest fold change at graph to the left).

TABLE 37: Indel Producing Variants of SEQ ID NO: 1

PL#####	Amino Acid Substitution Relative to SEQ ID NO: 1
PL26515	N286K
PL26567	E225K
PL26563	I80K
PL26562	S209F
PL26521	Y315M
PL26499	N193K
PL26517	M298L
PL26542	M295W
PL26519	A306K
PL26566	A218K
PL26549	K58W

Example 24: Indel activity of CasPhi.12 variants

[751] Variants of CasPhi.12 were generated and tested to identify variants with increased binding affinity and greater genomic editing efficiency. Briefly, plasmid constructs encoding CasPhi.12 variants were generated by mutating nucleotides that encode single amino acids of interest within regions that interact with guide nucleic acid or the target nucleic acid from the wild-type residue to arginine. Generated variants had a single amino acid alteration - an arginine (R) - at amino acid positions: 23, 24, 25, 26, 28, 29, 51, 52, 53, 54, 55, 56, 57, 125, 126, 127, 128, 129, 130, 131, 316, 511, 512, 513, 514, 515, 516, 517, 540, 541, 542, 543, 544, 545, 546, 590, 591, 592, 593, 594, 595, 596, 602, 603, 604,

605, 606, 607, or 608 (positions as identified with respect to **SEQ ID NO: 6**). Wild-type CasPhi.12 (WT) (**SEQ ID NO: 6**) and a non- targeting control (NTC) were included as controls.

[752] HEK293T cells were prepared, seeded and plated for plasmid lipofection. Plasmid preparations of the various constructs containing a guide targeting either B2M or FUT8 were incubated in reduced serum media (Opti-MEM) and lipid reagent. The mixture containing the CasPhi.12 variants plasmid constructs and guides targeting B2M or FUT8 were delivered by lipofection to HEK293T cells. 20 µL of the lipid:DNA mix and HEK293T cells were incubated for 3 days. The cells were lysed with QuickExtract (QE) solution and suspended for next-generation sequencing (NGS).

[753] Indels were detected by NGS at the targeted locis. Indel percentage was calculated as the fraction of sequencing reads containing insertions or deletions relative to an unedited reference sequence. Raw indel data is shown in **TABLE 38**, normalized data is shown in **TABLE 39**. To demonstrate relative nuclease activity, the mean of replicate values were plotted in relation to the two target loci and normalized to the wild type. Results can be seen in **FIG. 12**. Indel activity was highest in the compositions that contained the L26R mutation, which is projected to be in the region of the effector protein with T-strand and NT-strand PAM interacting domains (i.e., TPID, NPID). Without being bound by theory, it is contemplated that the L26R mutation improves the binding efficiency/strength of the CasPhi.12 variant with the target nucleic acid, *e.g.*, the genomic DNA.

TABLE 38: Raw indel data

	PL7889		**PL7895**	
Mutation	**FUT8**		**B2M**	
T23R	40.876993166	33.650773361	42.672132675	35.340833885
A24R	2.0717985355	1.8742526626	0.4884547069	0.5820182424
G25R	26.235242676	23.579793742	17.132684051	16.681480488
L26R	61.012715713	52.934638952	49.325521188	48.84152257
L28R	18.7297172	16.082247984	13.460261235	16.844182045
K29R	37.365774093	32.226375069	26.529411765	28.1023859
P51R	20.974950629	15.231072159	18.999797119	16.588401185
N52R	13.952373342	11.004470155	14.98967247	17.678765246
F53R	7.4508560558	3.7136329018	4.2908633217	3.3490401186
Q54R	10.056249143	9.4506364176	7.6043237808	6.5556894353
G55R	14.580841579	14.807440398	13.845069181	14.219509269
G56R	12.748262165	12.714336295	7.1302428256	6.0666392055
P57R	1.6268035247	1.427402026	1.1138014528	0.9756036626
L125R	22.820124834	19.15660432	20.540080799	20.479229975
I126R	2.5208217657	2.5047470122	1.3509364446	1.0429735518
I127R	1.0343822844	0.8854057611	0.6996770721	0.9276388107
K128R	25.039228872	0	18.465077373	19.070725503
N129R	8.4337349398	11.325185861	7.2006472492	7.1492134573
A130R	1.2349066959	1.6122961104	0.6531842733	0.9591373439
V131R	4.7246257382	6.2485207101	4.6503044842	4.91959273
T316R	4.1080912215	4.0438607493	1.2746585736	1.8757964976
Q511R	41.933318416	30.916100187	30.75384923	34.433360204
D512R	27.614669314	26.222434403	20.570071259	23.714054837
Y513R	2.8526315789	2.2555410691	1.0457860972	1.2256332438

	PL7889		PL7895	
Mutation	FUT8		B2M	
K514R	14.638888889	11.833137995	12	17.191358441
P515R	4.1509707331	4.3051913147	1.8143009605	3.8539001687
K516R	37.71281362	30.952255481	35.123797831	37.170249684
L517R	8.6568827791	9.1801985877	5.8177417853	6.1249518703
N540R	39.758416859	40.639304834	32.567012326	40.215294835
K541R	35.733512786	34.531846199	26.854219949	37.137085408
L542R	4.3135783946	3.2927431777	1.5424872607	2.3479778877
S543R	35.335650663	28.62076775	28.263608871	31.639555365
K544R	15.513361463	13.775012768	15.221751727	22.075256556
S545R	36.643797098	35.476798102	34.195879309	35.381935484
R546R	39.511115794	41.47524189	44.830917874	42.650810577
Y590R	17.935610181	19.827667477	16.536385632	18.375897231
K591R	26.49896623	24.826515262	27.340341134	24.878955495
P592R	22.166951656	21.95414915	25.254738687	27.131135213
K593R	18.929530895	15.483476132	21.108892282	20.287286144
K594R	37.837453822	35.152935923	39.927949565	39.558390986
E595R	0.6606606607	0.6674045008	0.3588248486	0.5730879145
N596R	7.9125392607	5.7279106944	5.614973262	7.8704643675
A602R	5.4086407661	5.4720236628	6.0059788024	4.7688306566
I603R	0.6626332469	0.2962231548	0.1692198076	0.1931174677
H604R	7.4059637498	6.4590818363	3.2286383145	3.590690372
K605R	42.978896104	31.594433113	36.626815532	36.827641895
A606R	6.1785747468	2.6755995562	2.9095074456	3.4120481928
L607R	1.0147234381	0.9888298846	0.3773219814	0.3559746007
T608R	11.564927858	10.79544249	12.954186414	13.533553201
WT	47.319725296	28.283171945	44.574162679	40.941168563
NTC	0.010448229	0.0080469944	0.0091675834	0.0229042602

TABLE 39: Normalized indel data

Mutations	PL7889 (Fut8)	PL7895 (B2M)
N129R	0.1678346281	0.2613613863
P515R	0.06629475	0.1118539953
S545R	0.8137756116	0.953976127
N596R	0.1577244167	0.1804292322
T23R	0.9124323574	0.9858170176
F53R	0.0893555958	0.1476784254
K128R	0.4390152383	0.3312067311
A130R	0.0188575628	0.0376614128
K516R	0.8455444154	0.9082681098
A602R	0.1260211633	0.1439241327
A24R	0.0125201515	0.0521964444
Q54R	0.1656141897	0.2580275868
WT	1.0001793128	1.0000383233
V131R	0.1119286224	0.1451474398
L517R	0.1396806276	0.2359402297
Y590R	0.4083307937	0.4995142547
I603R	0.0042378629	0.0126832857
G25R	0.3954873045	0.6589290532

Mutations	PL7889 (Fut8)	PL7895 (B2M)
G55R	0.3282406836	0.3887338886
NTC	0.0003751093	0.0002446458
T316R	0.0368474277	0.1078300525
N540R	0.8512550545	1.0634619272
K591R	0.6107520074	0.6789084853
H604R	0.0797582303	0.1834000739
L26R	1.1481525586	1.5072401411
G56R	0.1543494974	0.3368068579
Q511R	0.7624235022	0.9636166482
K541R	0.74843632	0.9294359654
P592R	0.6127002795	0.5836124445
K605R	0.8591164611	0.9864196986
L28R	0.3544379331	0.4604757299
P57R	0.0244374867	0.0403995443
D512R	0.5179429953	0.7121310015
L542R	0.0455025164	0.1006127192
K593R	0.4841658295	0.4551985056
A606R	0.0739363233	0.1171187077
K29R	0.6389683937	0.9205310736
L125R	0.4797580208	0.5552477401
Y513R	0.0265663081	0.0675684213
S543R	0.7006218039	0.8459843705
K594R	0.9296648018	0.9654813458
L607R	0.0085765682	0.0265020281
P51R	0.4162362375	0.4789156453
I126R	0.0279989473	0.0664757775
K514R	0.341419397	0.3501590858
K544R	0.4362223191	0.3874123575
E595R	0.0108995645	0.0175669995
T608R	0.3097981242	0.2957720946
N52R	0.3820869908	0.3301169775
I127R	0.0190329343	0.0253940218

Example 25: Indel activity of CasPhi.12 variants

[754] Variants of CasPhi.12 were generated and tested to identify variants with increased binding affinity and greater genomic editing efficiency. Briefly, plasmid constructs encoding CasPhi.12 variants were generated by mutating nucleotides that encode single amino acids of interest within regions that interact with guide nucleic acid or the target nucleic acid from the wild-type residue to arginine. Generated variants had a single amino acid mutation - an arginine (R) - at amino acid positions: 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 132, 133, 134, 135, 136, 137, 138, 139, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, or 210 (positions as identified with respect to **SEQ ID NO: 6**). Wild-type CasPhi.12 (WT) (**SEQ ID NO: 6**) was included as a baseline control.

[755] HEK293T cells were prepared, seeded and plated for plasmid lipofection. Plasmid preparations of the various constructs containing a guide targeting FUT8 were incubated in reduced serum media (Opti-MEM) and lipid reagent. The mixture containing the CasPhi.12 variants plasmid constructs and guides targeting FUT8 were delivered by lipofection to HEK293T cells. 20 µL of the lipid:DNA mix and HEK293T cells were incubated for 3 days. The cells were lysed with QuickExtract (QE) solution and suspended for next-generation sequencing (NGS).

[756] Indels were detected by NGS at the targeted loci. Indel percentage was calculated as the fraction of sequencing reads containing insertions or deletions relative to an unedited reference sequence. Raw indel data is shown in **TABLE 40** and **FIG. 13A**, normalized data against the WT baseline is shown in **TABLE 41** and **FIG. 13B**. To demonstrate relative nuclease activity, the mean of replicate values were normalized to the wild type. Indel activity was highest for the compositions that contained the E109R, H208R, K184R, K38R, L182R, Q183R, S108R, S198R, and the T114R mutation.

TABLE 40: Raw indel data

Mutation	Rep 1	Rep 2
G13R	4.5003309067	4.1178947368
F14R	12.206251426	10.393336418
K15R	27.370356749	22.535314891
L16R	1.0535099614	0.802580258
I17R	14.719523935	9.5904330155
R18R	34.806727129	25.182765593
N19R	16.213097413	13.097593142
H20R	19.400196657	15.205479452
S21R	18.029395754	16.194899658
R22R	1.5044870667	1.1980500702
N30R	26.281142394	24.000960999
E31R	15.780542986	13.319206902
G32R	2.5	1.5377024916
E33R	21.770462633	26.105442177
E34R	25.064671101	29.625677959
A35R	15.547151196	18.783196523
C36R	0.696969697	1.1538461538
K37R	31.155411752	31.969653318
K38R	31.628910464	32.978345758
F39R	1.250899928	1.6980606848
V40R	0.6051257712	0.5850757721
R41R	28.826509426	33.197420823
E42R	11.733734487	17.358958462
N43R	27.927927928	30.18134715
E44R	4.1102077687	3.1665758967
S108R	33.388189739	31.579476861
E109R	34.958382878	33.974801519
H110R	21.974554707	22.439066701
G111R	23.984652817	23.048262716
L112R	3.7715086034	3.8577358183
D113R	20.668558255	20.897323387
T114R	35.923423423	34.967500275

Mutation	Rep 1	Rep 2
V115R	10.044712488	10.327837787
P116R	16.302521008	16.266944734
Y117R	2.0328599276	2.40853098
K118R	20.304849885	20.082081895
E119R	9.8519562919	8.8972538284
A120R	0.355073818	0.1931034483
A121R	0.3860091382	0.3927986907
G122R	0.5516867143	0.7775524003
L123R	0.5246589717	0.4078111521
N124R	0.102541566	0.1495827429
N132R	23.942687747	24.2152872
T133R	5.033492823	6.9926526476
Y134R	3.541472507	3.9736914223
K135R	31.407035176	26.978154071
G136R	1.2723762724	1.2704005935
V137R	9.1770084912	9.3946376204
Q138R	19.021393773	17.279305355
V139R	26.953579359	26.877682403
G179R	7.2240685674	7.587201901
Y180R	31.608056794	29.996979156
L181R	3.7063059898	3.1621926853
L182R	33.270998909	32.022787451
Q183R	36.78236193	35.55438226
K184R	40.511697907	42.926622965
P185R	9.5792946999	9.2819310584
S186R	34.958563536	29.583292286
P187R	11.816121722	10.857672694
N188R	0.9019700926	0.7430664574
K189R	26.65433621	22.426631625
S190R	3.9630413002	3.0875918435
I191R	4.2350247997	3.8957915832
Y192R	1.5611092092	1.5761376155
C193R	5.5651643097	4.2355155825
Y194R	5.4571293226	5.7247150997
Q195R	10.986594051	12.267358626
S196R	18.595825427	22.389335156
V197R	6.8554396423	6.8561382948
S198R	33.756652153	32.667512154
P199R	1.9747615837	2.0315423684
K200R	22.593167702	31.040652023
P201R	5.6712672522	4.4456404736
F202R	0.1279426817	0.1712946268
I203R	19.580975077	18.28257717
T204R	0.1446712747	0.0895977063
S205R	24.3180736	21.372446307
K206R	24.267158153	23.54387292
Y207R	17.295995945	18.674256335
H208R	34.12089598	32.185903984
N209R	29.793089389	29.590754877
V210R	8.0384900074	6.7905547017
L26R	40.346760527	35.032894737
WT	28.693106202	31.212074063

TABLE 41: Normalized Data

Mutation	Normalized to WT
G13R	0.1438644472
F14R	0.3772559859
K15R	0.8330777308
L16R	0.03098380159
I17R	0.4058072581
R18R	1.001407432
N19R	0.4892847401
H20R	0.5776741837
S21R	0.5713077777
R22R	0.04511357991
N30R	0.8393615238
E31R	0.4857634976
G32R	0.06740155816
E33R	0.7991947374
E34R	0.9129485767
A35R	0.5730781139
C36R	0.0308957563
K37R	1.053749689
K38R	1.078491976
F39R	0.04922713862
V40R	0.01986809051
R41R	1.035368393
E42R	0.4856456958
N43R	0.97002087
E44R	0.121471693
S108R	1.084508323
E109R	1.150704899
H110R	0.7413986772
G111R	0.7851226776
L112R	0.127355337
D113R	0.693861223
T114R	1.183385533
V115R	0.3400799428
P116R	0.543683628
Y117R	0.07414034793
K118R	0.6741809573
E119R	0.312981449
A120R	0.009150748965
A121R	0.01300067583
G122R	0.02218905124
L123R	0.01556576776
N124R	0.004208722982
N132R	0.8039033475
T133R	0.2007530136
Y134R	0.125450986
K135R	0.9746267182
G136R	0.04244669417
V137R	0.3100173646

Mutation	Normalized to WT
Q138R	0.6059692829
V139R	0.8986077919
G179R	0.2472452366
Y180R	1.028375771
L181R	0.1146561724
L182R	1.089952256
Q183R	1.207520683
K184R	1.392839826
P185R	0.3148513313
S186R	1.077400244
P187R	0.3784947198
N188R	0.02746067273
K189R	0.8193109113
S190R	0.1176965517
I191R	0.1357281014
Y192R	0.05237020923
C193R	0.1636032118
Y194R	0.1866590564
Q195R	0.388179329
S196R	0.684167219
V197R	0.2288880173
S198R	1.108821708
P199R	0.06687742086
K200R	0.8953118827
P201R	0.1688820179
F202R	0.004995182506
I203R	0.6320580638
T204R	0.003910663151
S205R	0.7627140041
K206R	0.798111797
Y207R	0.6004531181
H208R	1.10686254
N209R	0.9912973136
V210R	0.2475419428
L26R	1.258316141
WT	1

Example 26: *Cis* cleavage assay for CasPhi.12 variants

[757] This experiment was performed to determine if variants of CasPhi.12 can perform *cis* cleavage. Briefly, effector protein, CasPhi.12, was mutated to generate nuclease dead variants using the following substitutions: D369A (**SEQ ID NO: 7**), D369N (**SEQ ID NO: 8**), E567A (**SEQ ID NO: 9**), E567Q (**SEQ ID NO: 10**), D658A (**SEQ ID NO: 11**), and D658N (**SEQ ID NO: 12**). The variants were then complexed with crRNA having nucleotide sequence of CUUUCAAGACUAAUAGAUUGCUCCUUACGAGGAGACUAUUAAAUACUCGUAUUGCU (**SEQ ID NO: 1847**) for 20 minutes at room temperature to form RNP complexes. The RNP complexes were added to an IVE reaction mix. *Cis* cleavage assay was carried out with 5 µl of RNP for at least 30 minutes at 37 °C for identifying catalytically dead variants. A plasmid containing TTTG PAM was used

as target nucleotide. Wildtype CasPhi.12 (CasPhi_WT) (**SEQ ID NO: 6**) with supercoiled plasmid DNA was included as a positive control. *Cis* cleavage by each complex was assessed by gel electrophoresis. The results are shown in **FIG. 14**. An absence of a linearized product indicates a lack of cleavage activity. These results indicate that none of the variants assayed showed detectable *cis* cleavage activity.

[758] Fluorescence polarization assays were performed for the RNP complexes comprising CasPhi.12 variants, D369A (**SEQ ID NO: 7**), E567A (**SEQ ID NO: 9**), E567Q (**SEQ ID NO: 10**), and D658N (**SEQ ID NO: 12**), and crRNA (**SEQ ID NO: 1847**), the results of which showed that the RNP complex were able to bind the DNA.

Example 27: *Cis* cleavage assay for CasM.265466 variants

[759] This experiment was performed to determine if variants of CasM.265466 can perform *cis* cleavage. Briefly, effector protein, CasM.265466 (**SEQ ID NO: 231**), was mutated to generate nuclease dead variants using the following substitutions: D237A, D237A & D418A, D418A, D418N, E335A, and E335Q. The variants were then complexed with sgRNA (**SEQ ID NO: 1828**) for 20 minutes at room temperature. The complexes were added to an IVE reaction mix. *Cis* cleavage assay was carried out with 5 µl of RNP for at least 30 minutes at 37 °C for identifying catalytically dead variants. A plasmid containing TTTG PAM was used as target nucleotide. Wildtype CasM.265466 (Cas466_WT); CasM.265466 variant D220R (Cas466_D220R); dAsCas12a and Cas14a1_D326A/D510A (dCas14a.1) were included as controls. *Cis* cleavage by each complex was assessed by gel electrophoresis. The results are shown in **FIG. 15**. An absence of a linearized product indicates a lack of cleavage activity.

[760] Fluorescence polarization assays were performed for the RNP complexes comprising CasM.265466 variants, D237A, D237A & D418A, D418A, D418N, E335A, and E335Q, and sgRNA (**SEQ ID NO: 1828**), the results of which showed that the RNP complex were able to bind the DNA.

[761] While preferred embodiments of the present disclosure have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the disclosure. It should be understood that various alternatives to the embodiments of the disclosure described herein can be employed in practicing the disclosure. It is intended that the following claims define the scope of the disclosure and that methods and structures within the scope of these claims and their equivalents be covered thereby.

Example 28: Indel introduction using CasPhi.32 compositions in primary human hepatocytes obtained from multiple donors.

[762] CasPhi.32 system is tested for it indel activity of cleaving within intron 1 of human albumin gene in primary human hepatocytes that were obtained from multiple donors. The CasPhi.32 system including the effector protein CasPhi.32 (**SEQ ID NO: 47**) and any one of each of the gRNAs described

in **TABLE 9** are used for cleaving within intron 1 of human albumin gene (**SEQ ID NO: 266**) in primary human hepatocytes obtained from the different donors. SpyCas9 system is used as a positive control.

[763] For determining indel activity, 100,000 primary human hepatocytes are seeded in 96-well plates. After 24 hours, cells are transfected with 2 μg total RNA (1:1 mRNA to gRNA ratio) using Lipofectamine MessengerMax. 48 hours post-transfection, DNA is extracted, and gene editing is quantified by amplicon sequencing.

Example 29: CasPhi.32 mediated integration of a bidirectional AAV reporter in eukaryotic cells

[764] A Cas12j32 system is tested for its ability to facilitate homology-independent targeted insertion (HITI) of a donor nucleic acid encoding luciferase protein into the intron 1 of human albumin gene in Hep3B cells and primary human hepatocyte cells. Briefly, 20,000 cells are seeded in a 96-well plate and transfected using lipofectamine MessengerMax. Hep3B cells are transfected with 200 ng total RNA (1:1 mRNA to gRNA ratio), whereas primary human hepatocyte cells are transfected with 500 ng RNA (1:1 mRNA to gRNA ratio) The gRNA is any one of the sequences recited in **TABLE 9**. One or more SpyCas9 gRNA sequences are used as positive controls. The transfection is carried in conjunction with AAV transduction, where AAV carrying reporter AAV vector is added to the media. The reporter AAV vector is packaged with the donor nucleic acid encoding a luciferase protein. The donor nucleic acid is co-delivered to the cells with the nuclease mRNA and gRNA. The cells are harvested, and the media is collected from three days to seven days post-transfection and transduction. The media is subject to a luciferase assay using the Promega Nano-Glo system. The cells are subject to DNA extraction, PCR amplified and sequenced via NGS sequencing analysis.

Example 30: Effector protein mediated integration of a bidirectional AAV reporter in eukaryotic cells

Example 30.1: Dose optimization of effector protein mediated integration

[765] Effector protein systems (*e.g.,* CasPhi.12 L26R, CasM.265466,) were tested for their ability to generate indel in primary human hepatocytes within or adjacent to intron 1 of human albumin gene as described in at least Examples 1, 2, 3, and 2 was tested. In brief, the effector protein system (*e.g.,* CasPhi.12 L26R, CasM.265466) was tested for its ability to facilitate homology-independent targeted insertion (HITI) of a donor nucleic acid encoding a luciferase protein (*e.g.,* nano luciferase reporter, nanoLuc, nLuc, nLuc AAV reporter) into the intron 1 of human albumin gene in primary human hepatocyte cells. 20,000 cells were seeded in a low adhesion 96-well plate and transfected using lipofectamine MessengerMax. Primary human hepatocyte cells were transfected with different concentrations of RNA (*e.g.,* 25 ng, 100 ng, 400 ng) (1:1 mRNA to gRNA ratio). The gRNA used for CasPhi.12 L26R system included **SEQ ID NOS: 652, 653, 685, 690, 705, 706, 713, 714, 727, 774, 777, 785, 790, 808** and **809**. The gRNA used for CasM.265466 system included **SEQ ID NOS: 856, 858,**

862, 864, and **866-868**. One or more SpyCas9 gRNA sequences were used as positive controls. The transfection was carried in conjunction with AAV transduction such that the different concentrations of RNA were paired with different concentrations of viral particles or MOI (multiplicity of infection) (*e.g.*, $2.5x10^3$, $1x10^4$, $4x10^4$), where AAV carrying reporter AAV vector was added to the media. The reporter AAV vector was packaged with the donor nucleic acid encoding a luciferase protein (*e.g.*, nLuc). The donor nucleic acid was co-delivered to the cells with the nuclease mRNA and gRNA. The composition comprising the effector protein system was incubated for 2 hours while rocking the low adhesion 96-well plate before being transferred to a Collagen I coated 96-well plate for culture. 48 hours post-transfer, the cells were harvested and were subject to DNA extraction. Indels were quantified by NGS and integration assessed by luciferase assay.

[766] **FIG. 16A** shows % indel activity of the effector protein system (*e.g.*, CasPhi.12 L26R, CasM.265466) within or adjacent to intron 1 of human albumin gene as related to different concentrations of RNA and MOI. As shown in **FIG. 16A,** the effector protein system (*e.g.*, CasPhi.12 L26R, CasM.265466) may be used to generate indels within intron 1 of human albumin gene, as compared to the positive control.

[767] **FIG. 16B** shows relative light units (RLU) as a measure of integration activity of the effector protein system (*e.g.*, CasPhi.12 L26R, CasM.265466) within or adjacent to intron 1 of human albumin gene as related to different concentrations of RNA and MOI. As shown in **FIG. 16B,** the effector protein system (*e.g.*, CasPhi.12 L26R, CasM.265466) may be used to insert a nucleotide sequence encoding a donor nucleic acid encoding a luciferase protein (*e.g.*, nLuc) within intron 1 of human albumin gene, as compared to the positive control. In particular, CasPhi.12 L26R and CasM.265466 can achieve 50%-70% integration. Also as shown in **FIG. 16B,** the highest levels of integration were achieved upon pairing the effector protein system with 400 ng of RNA and MOI of $1x10^4$.

Example 30.2: Dose optimization of effector protein mediated integration

[768] Similar to Example 33.1, an effector protein (*e.g.*, CasPhi.12 L26R, CasM.265466) system capable of indel activity within or adjacent to intron 1 of human albumin gene was tested for its ability to facilitate homology-independent targeted insertion (HITI) of a donor nucleic acid encoding a luciferase protein (*e.g.*, nLuc, nLuc AAV reporter) into the intron 1 of human albumin gene in primary human hepatocyte cells. Briefly, 20,000 cells were seeded in a low adhesion 96-well plate and transfected using lipofectamine MessengerMax. Primary human hepatocyte cells from two different human donors were transfected with 400 ng of RNA. The gRNA used for CasPhi.12 L26R system included **SEQ ID NOS: 652, 653, 685, 690, 705, 706, 713, 714, 727, 774, 777, 785, 790, 808** and **809**. The gRNA used for CasM.265466 system included **SEQ ID NOS: 856, 858, 862, 864,** and **866-868**. One or more SpyCas9 gRNA sequences were used as positive controls. The transfection was carried in conjunction with AAV transduction with MOI of $1x10^4$, where AAV carrying reporter AAV vector was

added to the media. The reporter AAV vector was packaged with the donor nucleic acid encoding a luciferase protein (*e.g.*, nLuc). The donor nucleic acid was co-delivered to the cells with the nuclease mRNA and gRNA. The composition comprising the effector protein system was incubated for 2 hours while rocking the low adhesion 96-well plate before being transferred to a Collagen I coated 96-well plate for culture. 48 hours post-transfer, the cells were harvested and were subject to DNA extraction. Indels were quantified by NGS and integration assessed by luciferase assay and reverse transcription droplet digital PCR (RT-ddPCR) as measured via RT-ddPCR.

[769] **FIG. 17** shows % integration products as a measure of integration activity of the effector protein system (*e.g.*, CasPhi.12 L26R, CasM.265466) within or adjacent to intron 1 of human albumin gene in primary human hepatocyte cells from two different human donors as related to different concentrations of RNA and MOI. As shown in **FIG. 17,** the effector protein system (*e.g.*, CasPhi.12 L26R, CasM.265466) may be used to insert a nucleotide sequence encoding a donor nucleic acid encoding a luciferase protein (*e.g.*, nLuc) within intron 1 of human albumin gene, as compared to the positive control. For example, CasPhi.12 L26R and CasM.265466 can achieve 3% and 4% integration, respectively.

[770] While preferred embodiments of the present disclosure have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the disclosure. It should be understood that various alternatives to the embodiments of the disclosure described herein can be employed in practicing the disclosure. It is intended that the following claims define the scope of the disclosure and that methods and structures within the scope of these claims and their equivalents be covered thereby.

CLAIMS

What is claimed is:

1. A composition comprising:

   i. an effector protein, or a nucleic acid encoding the effector protein;

   ii. a guide nucleic acid, or a nucleic acid encoding the guide nucleic acid; and

   iii. a donor nucleic acid,

   wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**,

   wherein the guide nucleic acid comprises a spacer sequence that hybridizes to a target sequence in a target nucleic acid,

   wherein the target sequence comprises a nucleotide sequence within a human safe harbor locus, and

   wherein the donor nucleic acid encodes a transgene that comprises a functional human protein that is expressed upon incorporation into the human safe harbor locus.

2. The composition of claim 1, wherein the target sequence is in proximity to or adjacent to a protospacer adjacent motif (PAM) sequence comprising any one of the sequences recited in **TABLE 3**.

3. The composition of claim 1 or 2, wherein the human safe harbor locus is an intron.

4. The composition of any one of claims 1-3, wherein the human safe harbor locus comprises any one of the sequences recited **TABLE 10**.

5. The composition of any one of claims 1-4, wherein the guide nucleic acid comprises a nucleotide sequence that is at least 90% identical to any one of the sequences recited in **TABLE 4**, **TABLE 5**, **TABLE 6**, **TABLE 7**, **TABLE 9,** and any combination thereof.

6. The composition of any one of claims 1-5, wherein the guide nucleic acid comprises a spacer sequence that is at least 90% identical to any one of the sequences recited in **TABLE 4** and **TABLE 5**.

7. The composition of any one of claims 1-6, wherein the target sequence comprises a nucleotide sequence with at least 90% sequence identity to any one of sequences recited in **TABLE 10**.

8. The composition of any one of claims 1-7, wherein the target sequence comprises a nucleotide sequence within **SEQ ID NO: 266**.

9. The composition of any one of claims 1-8, wherein the spacer sequence is at least 90% identical to any one of sequences recited in **TABLE 4**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6, 13, 47**, and **228-231**.

10. The composition of any one of claims 1-9, wherein the guide nucleic acid comprises a repeat sequence of **SEQ ID NO: 252** or **1848**, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6** and **228-230**, and wherein the target sequence is adjacent to a PAM of **SEQ ID NO: 234-237**.
11. The composition of any one of claims 1-9, wherein the guide nucleic acid comprises a repeat sequence of **SEQ ID NO: 253**, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 13**, and wherein the target sequence is adjacent to a PAM of **SEQ ID NO: 238, 240-247**.
12. The composition of any one of claims 1-9, wherein the guide nucleic acid comprises a handle sequence of **SEQ ID NO: 259-261**, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 13**, and wherein the target sequence is adjacent to a PAM of **SEQ ID NO: 238, 240-247**.
13. The composition of any one of claims 1-9, wherein the guide nucleic acid comprises a repeat sequence of **SEQ ID NO: 254**, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 47**, and wherein the target sequence is adjacent to a PAM of **SEQ ID NO: 241-244**.
14. The composition of any one of claims 1-9, wherein the guide nucleic acid comprises a repeat sequence of **SEQ ID NO: 255** or **1789**, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 231**, and wherein the target sequence is adjacent to a PAM of any one of **SEQ ID NO: 245-248, 1829,** and **1834-1844**.
15. The composition of any one of claims 1-9, wherein the guide nucleic acid comprises any one of handle sequences of **SEQ ID NO: 262-264**, wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 231**, and wherein the target sequence is adjacent to a PAM of any one of **SEQ ID NO: 245-248, 1829,** and **1834-1844**.
16. The composition of any one of claims 1-9, wherein the guide nucleic acid comprises a spacer sequence that is at least 90% identical to any one of **SEQ ID NO: 348, 558, 401, 409, 473,** and **486**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6** and **228-230**.
17. The composition of any one of claims 1-9, wherein the guide nucleic acid comprises a spacer sequence that is at least 90% identical to any one of **SEQ ID NO: 592, 600,** and **602**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 231**.
18. The composition of any one of claims 1-9, wherein the guide nucleic acid comprises a nucleotide sequence that is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 9**.
19. The composition of any one of claims 1-9, wherein the guide nucleic acid comprises a nucleotide sequence that is at least 90% identical to any one of **SEQ ID NO: 652, 685, 705,**

713, 777, and **790**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6,** and **228-230**.

20. The composition of any one of claims 1-9, wherein the guide nucleic acid comprises a nucleotide sequence that is at least 90% identical to any one of **SEQ ID NO: 856, 864,** and **866**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 231**.

21. The composition of any one of claims 1-9, wherein the target sequence comprises a nucleotide sequence within **SEQ ID NO: 1795**.

22. The composition of any one of claims 1-9, wherein the spacer sequence is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 5**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of the amino acid sequences of **SEQ ID NO: 6, 13, 47,** and **228-231**.

23. The composition of any one of claims 1-22, comprising a nucleic acid expression vector, wherein the expression vector comprises at least one of the nucleic acid encoding the effector protein; the nucleic acid encoding the guide nucleic acid; and the donor nucleic acid.

24. The composition of claim 23, wherein the nucleic acid expression vector is a viral vector.

25. The composition of claim 24, wherein the viral vector is an adeno associated viral (AAV) vector.

26. The composition of any one of claims 1-22, wherein the nucleic acid encoding the effector protein is a messenger RNA.

27. The composition of any one of claims 1-26, comprising a lipid or a lipid nanoparticle.

28. A pharmaceutical composition, comprising the composition of any one of claims 1-27; and a pharmaceutically acceptable excipient.

29. A system comprising components for introduction of a donor nucleic acid encoding a functional human protein into a human safe harbor locus, wherein the components comprise:

i. an effector protein, or a nucleic acid encoding the effector protein;

ii. a guide nucleic acid, or a nucleic acid encoding the guide nucleic acid; and

iii. a donor nucleic acid,

wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of the amino acid sequences recited in **TABLE 1A** and **TABLE 1B**,

wherein the guide nucleic acid comprises a spacer sequence that hybridizes to a target sequence in a target nucleic acid,

wherein the target sequence comprises a nucleotide sequence within a human safe harbor locus, and

wherein the donor nucleic acid encodes a transgene that comprises a functional human protein that is expressed upon incorporation into the human safe harbor locus.

30. The system of claim 29, wherein the target sequence is in proximity to or adjacent to a protospacer adjacent motif (PAM) sequence comprising any one of the nucleotide sequences recited in **TABLE 3**.
31. The system of claim 29 or 30, wherein the human safe harbor locus is an intron.
32. The system of any one of claims 29-31, wherein the human safe harbor locus comprises any one of the nucleotide sequences recited **TABLE 10**.
33. The system of any one of claims 29-32, wherein the guide nucleic acid comprises a nucleotide sequence that is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 4**, **TABLE 5**, **TABLE 6**, **TABLE 7**, **TABLE 9,** and any combination thereof.
34. The system of any one of claims 29-33, wherein the guide nucleic acid comprises a spacer sequence that is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 4** and **TABLE 5**.
35. The system of any one of claims 29-34, wherein the target sequence comprises a nucleotide sequence with at least 90% sequence identity to any one of the nucleotide sequences recited in **TABLE 10**.
36. The system of any one of claims 29-35, wherein the target sequence comprises a nucleotide sequence within **SEQ ID NO: 266**.
37. The system of claim 36, wherein the spacer sequence is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 4**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6, 13, 47**, and **228-231**.
38. The system of claim 37, wherein the guide nucleic acid comprises a spacer sequence that is at least 90% identical to any one of **SEQ ID NO: 348, 558, 401, 409, 473,** and **486**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6** and **228-230**.
39. The system of claim 37, wherein the guide nucleic acid comprises a spacer sequence that is at least 90% identical to any one of **SEQ ID NO: 592, 600,** and **602**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 231**.
40. The system of claim 37, wherein the guide nucleic acid comprises a nucleotide sequence that is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 9**.
41. The system of claim 37, wherein the guide nucleic acid comprises a nucleotide sequence that is at least 90% identical to any one of **SEQ ID NO: 652, 685, 705, 713, 777,** and **790**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of **SEQ ID NO: 6** and **228-230**.
42. The system of claim 37, wherein the guide nucleic acid comprises a nucleotide sequence that is at least 90% identical to any one of **SEQ ID NO: 856, 864,** and **866**, and wherein the

effector protein comprises an amino acid sequence that is at least 90% identical to **SEQ ID NO: 231**.

43. The system of any one of claims 29-42, wherein the target sequence comprises a nucleotide sequence within **SEQ ID NO: 1795**.

44. The system of claim 43, wherein the spacer sequence is at least 90% identical to any one of the nucleotide sequences recited in **TABLE 5**, and wherein the effector protein comprises an amino acid sequence that is at least 90% identical to any one of the amino acid sequences of **SEQ ID NO: 6, 13, 47** and **228-231**.

45. The system of any one of claims 29-44, comprising a nucleic acid expression vector, wherein the expression vector comprises at least one of the nucleic acid encoding the effector protein; the nucleic acid encoding the guide nucleic acid; and the donor nucleic acid.

46. The system of claim 45, wherein the nucleic acid expression vector is a viral vector.

47. The system of claim 46, wherein the viral vector is an adeno associated viral (AAV) vector.

48. The system of any one of claims 29-44, wherein the nucleic acid encoding the effector protein is a messenger RNA.

49. The system of any one of claims 29-48, comprising a lipid or a lipid nanoparticle.

50. The system of any one of claims 29-49, wherein any two components of the system are provided in different solutions or containers.

51. A method of expressing a functional human protein in a cell, the method comprising contacting the cell with the composition of any one of claims 1-27, or the system of any one of claims 29-50.

52. A method of treating a disease associated with a mutation or aberrant expression of a human protein in a subject in need thereof, the method comprising administering to the subject the composition of any one of claims 1-27 or the components of the system of any one of claims 29-50.

53. The method of claim 52, wherein at least two components of the system are administered separately or simultaneously.

54. The method of claim 52 or 53, wherein the subject has cystic fibrosis, muscular dystrophy, Freidreich's ataxia, amylotrophic lateral scleorsis, hemophilia, Huntington's disease, retinal dystrophy, Rett syndrome, sickle cell disease, or a combination thereof.

55. The method of any one of claims 52-54, wherein administering increases the amount of the functional human protein in a biological sample of the subject relative to the amount or concentration of functional protein in the biological sample before administering.

56. The method of claim 55, wherein the amount or concentration of functional protein in the biological sample is increased by at least 10%, at least 20%, at least 30%, at least 40%, or at least 50%.

57. The method of claim 55 or 56, wherein the biological sample is selected from blood, serum, plasma, urine, saliva, and cerebrospinal fluid.

58. A cell comprising the composition of any one of claims 1-27.

59. A cell that comprises the target nucleic acid modified by the composition of any one of claims 1-27.

60. A cell modified by the composition of any one of claims 1-27, the system of claims 29-50, or the method of claim 51.

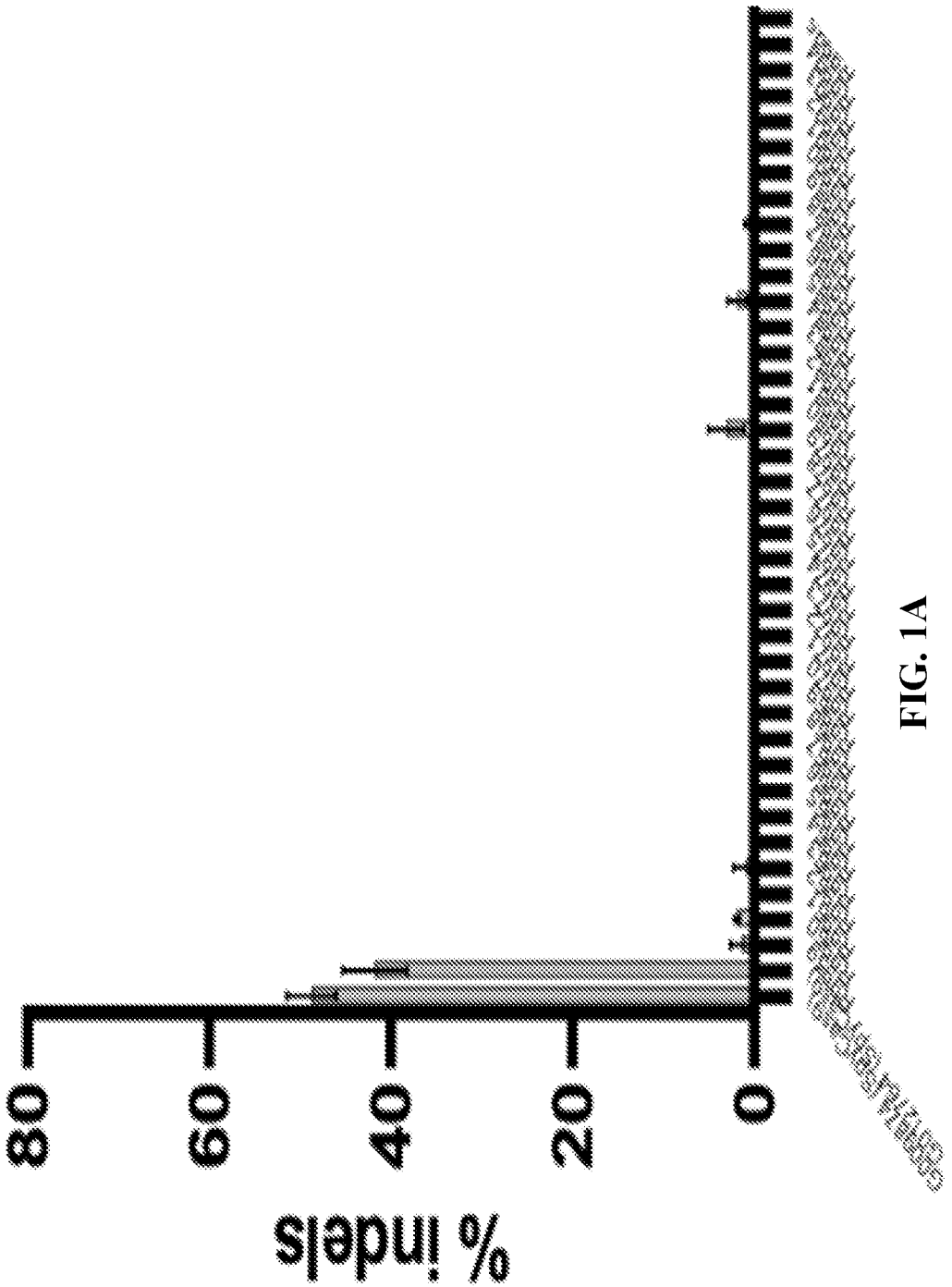

FIG. 1A

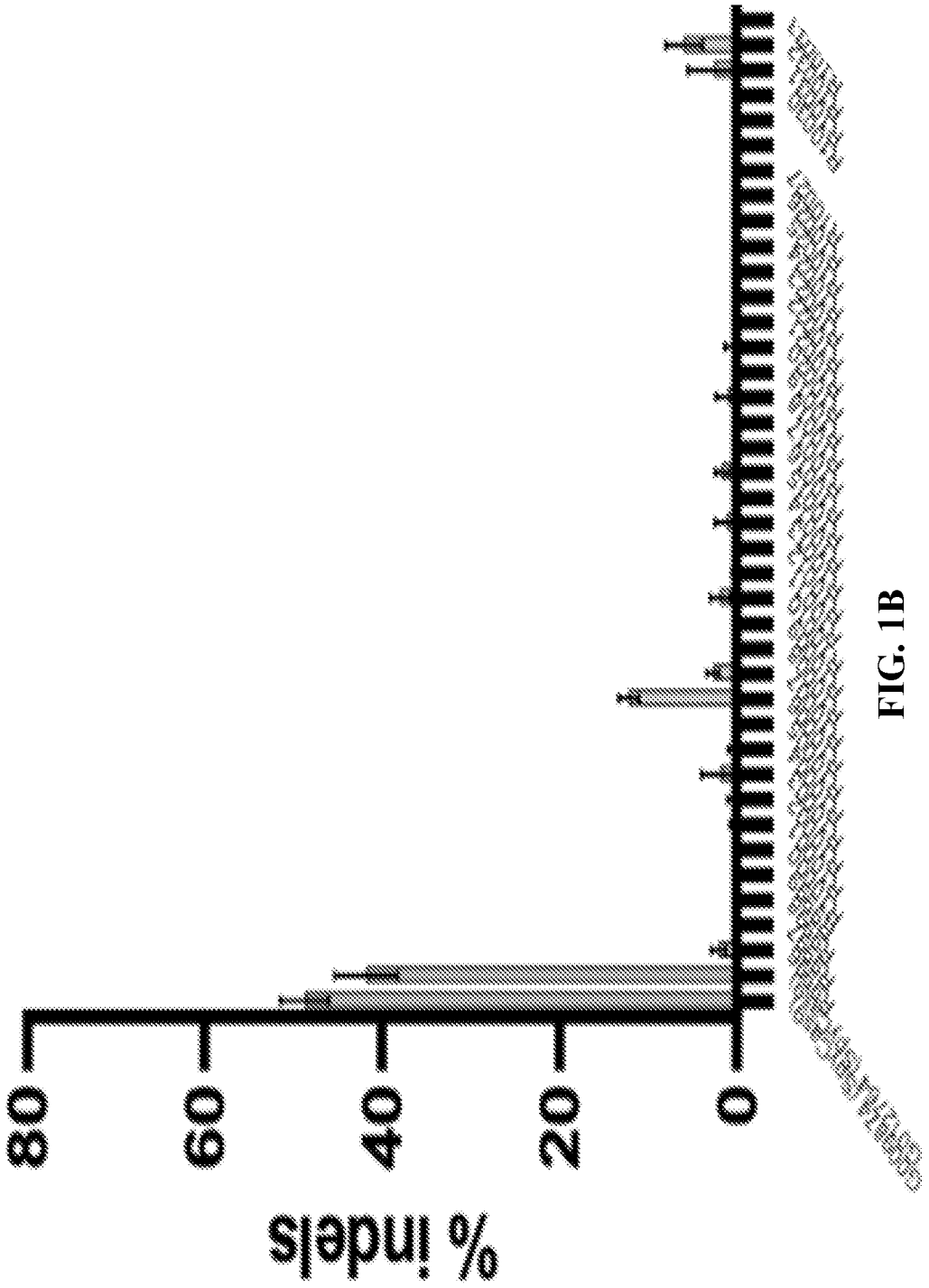

FIG. 1B

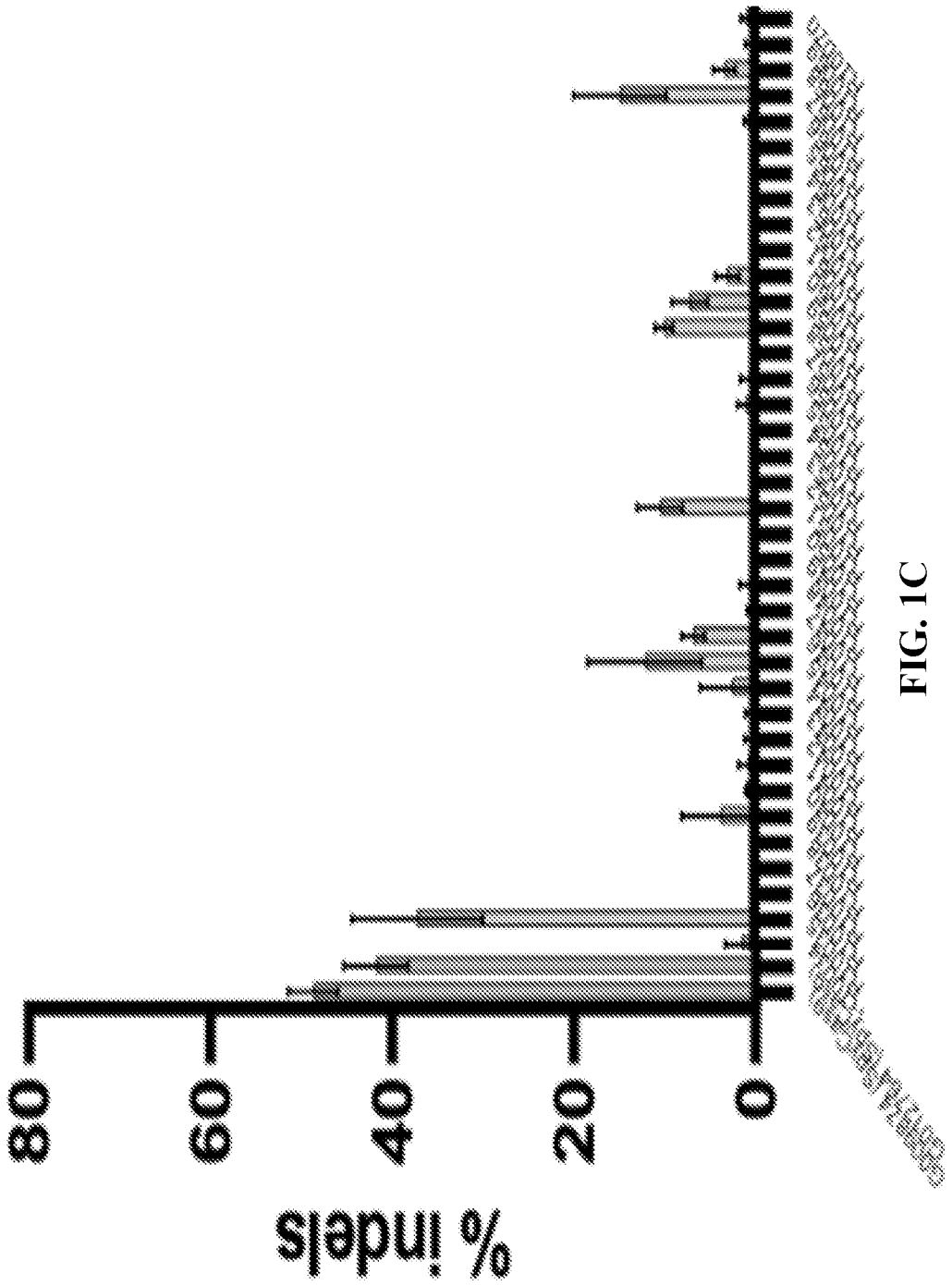

FIG. 1C

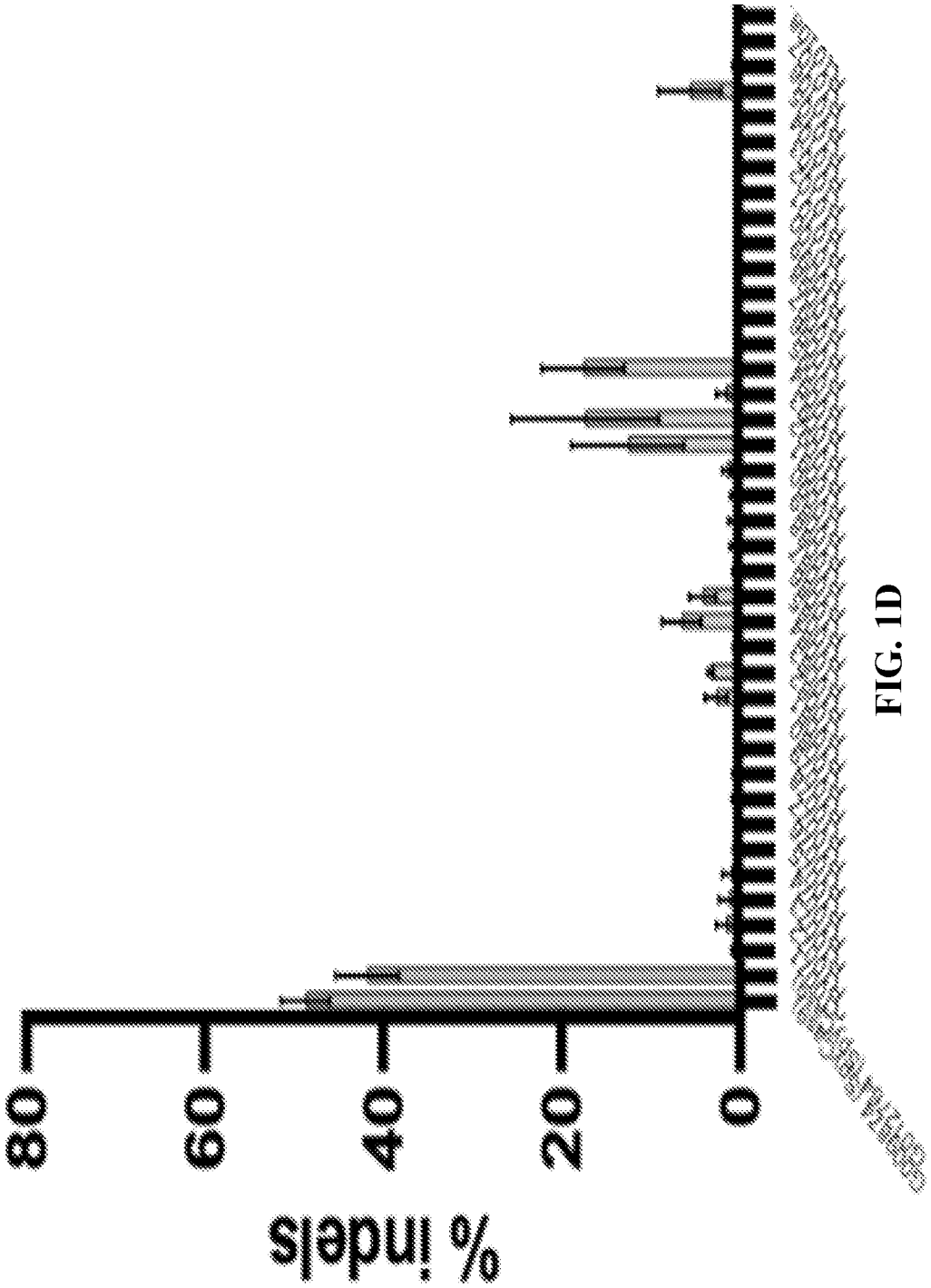

FIG. 1D

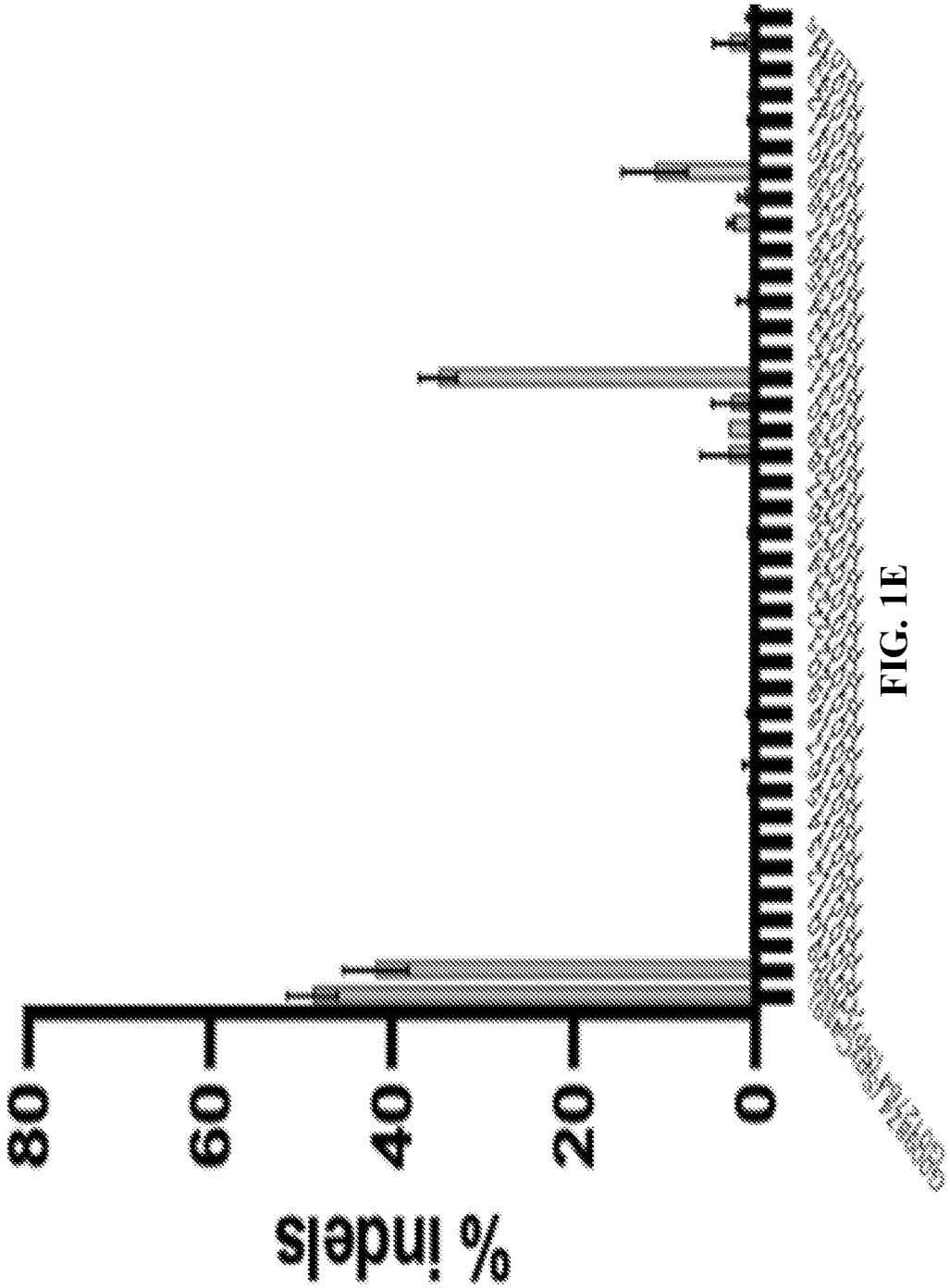

FIG. 1E

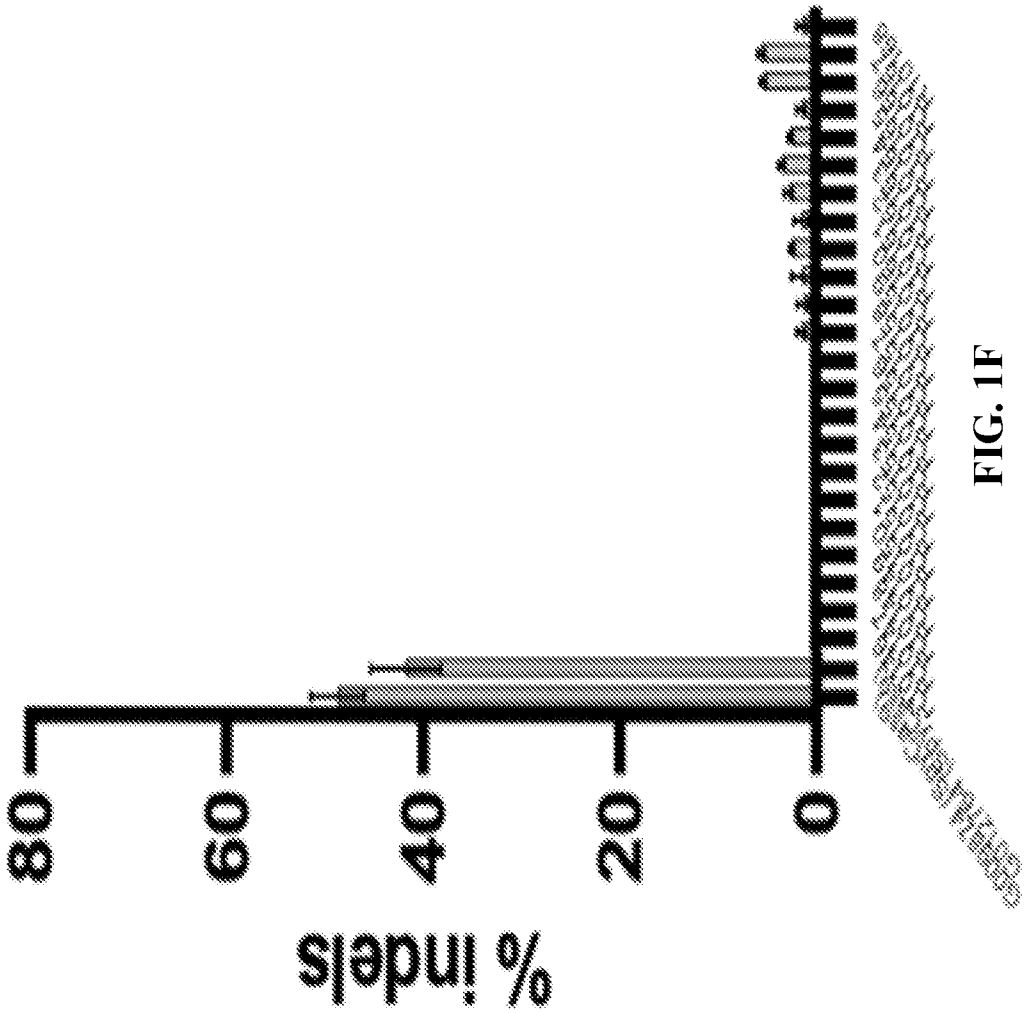

FIG. 1F

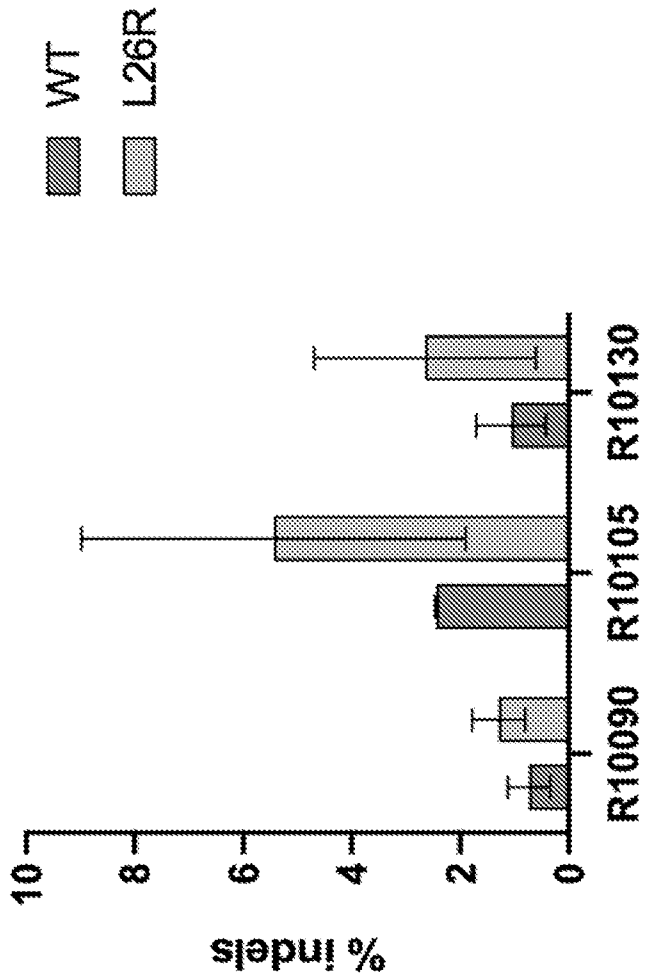

FIG. 2


% Indels
60
50
40
30
5
4
3
2
1
0
G009874 (SpyCas9)
G012764 (SpyCas9)
R10169
R10170
R10171
R10172
R10173
R10174
R10175
R10176
R10177
R10178
R10179
R10180
R10181
R10182
R10183
R10184
R10185
R10186
R10187
R10188
R10189
R10190
R10191
R10192
R10193

FIG. 3

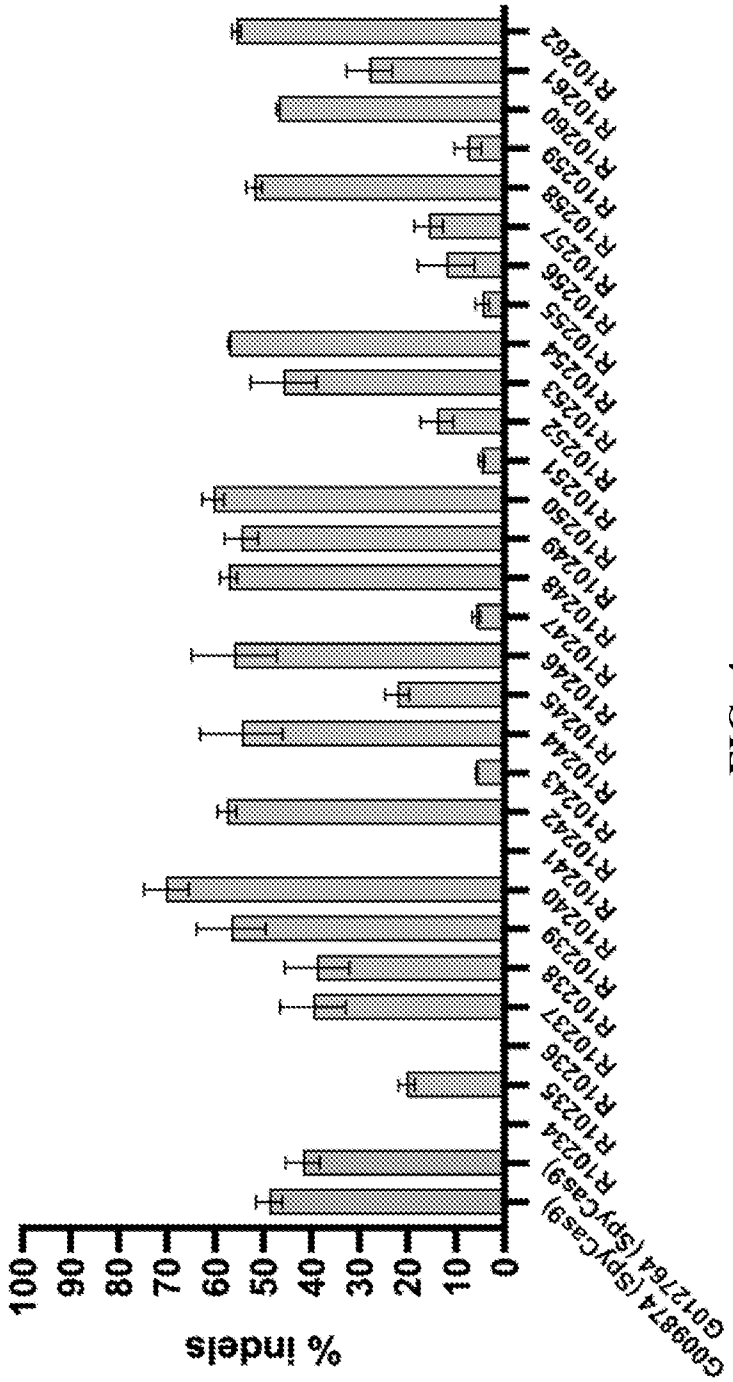
% Indels
100
90
80
70
60
50
40
30
20
10
0
G009874 (SpyCas9)
G012764 (SpyCas9)
R10234
R10235
R10236
R10237
R10238
R10239
R10240
R10241
R10242
R10243
R10244
R10245
R10246
R10247
R10248
R10249
R10250
R10251
R10252
R10253
R10254
R10255
R10256
R10257
R10258
R10259
R10260
R10261
R10262

FIG. 4

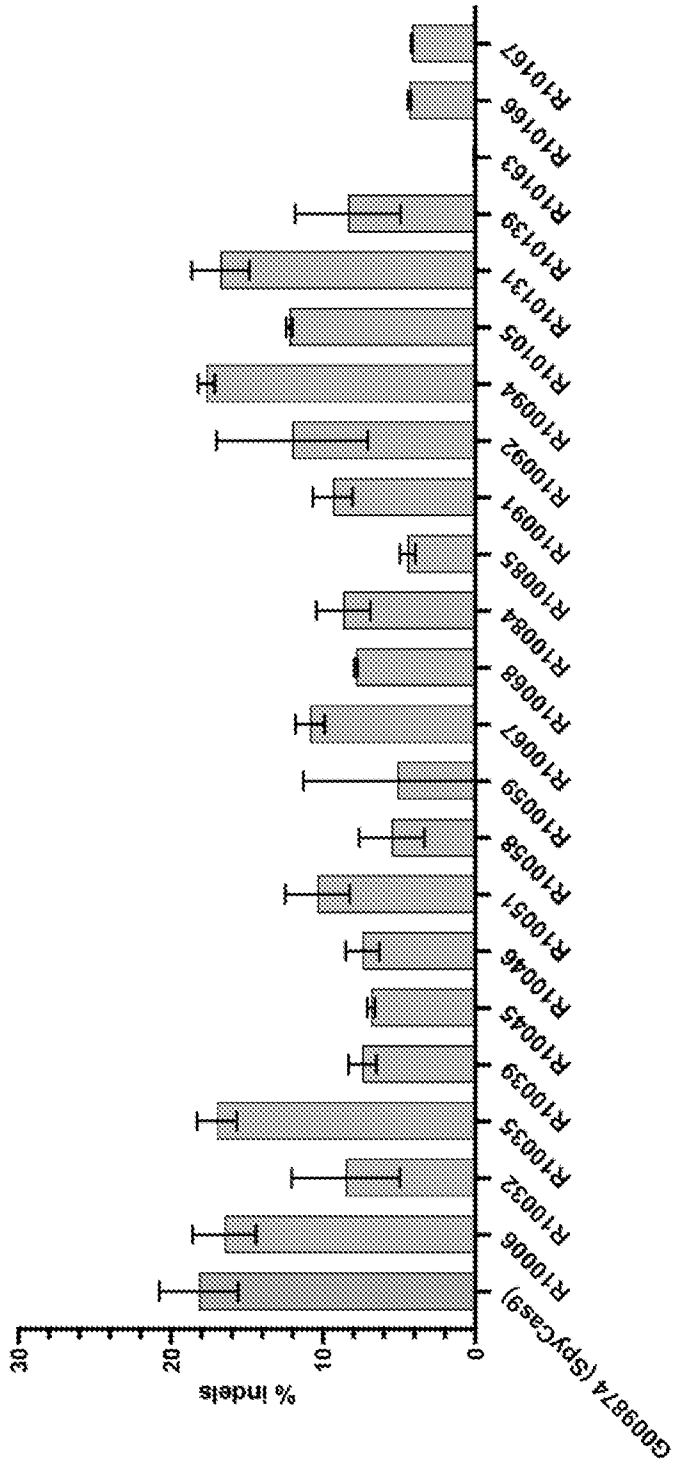
% indels
30
20
10
0
G009874 (SpyCas9)
R10006
R10032
R10035
R10039
R10045
R10046
R10051
R10058
R10059
R10067
R10068
R10084
R10085
R10091
R10092
R10094
R10105
R10131
R10139
R10163
R10166
R10167

FIG. 5

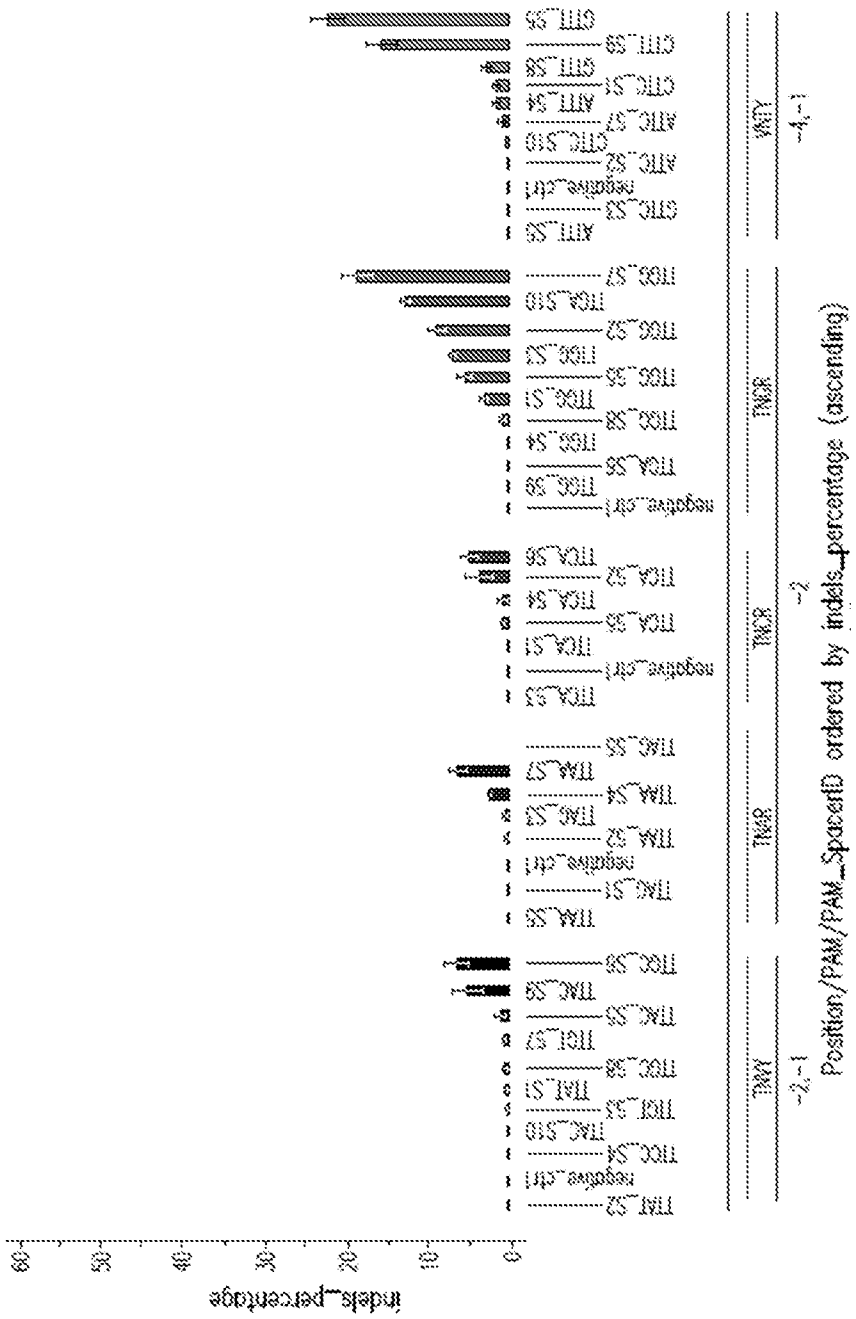
indels_percentage
0
10
20
30
40
50
60
TTAT
TTAR
TWCR
TWCR
RTAA
WATY
-2,-1
-2
-4,-1
Position/PAM/PAM_SpacerID ordered by indels_percentage (ascending)

FIG. 6A

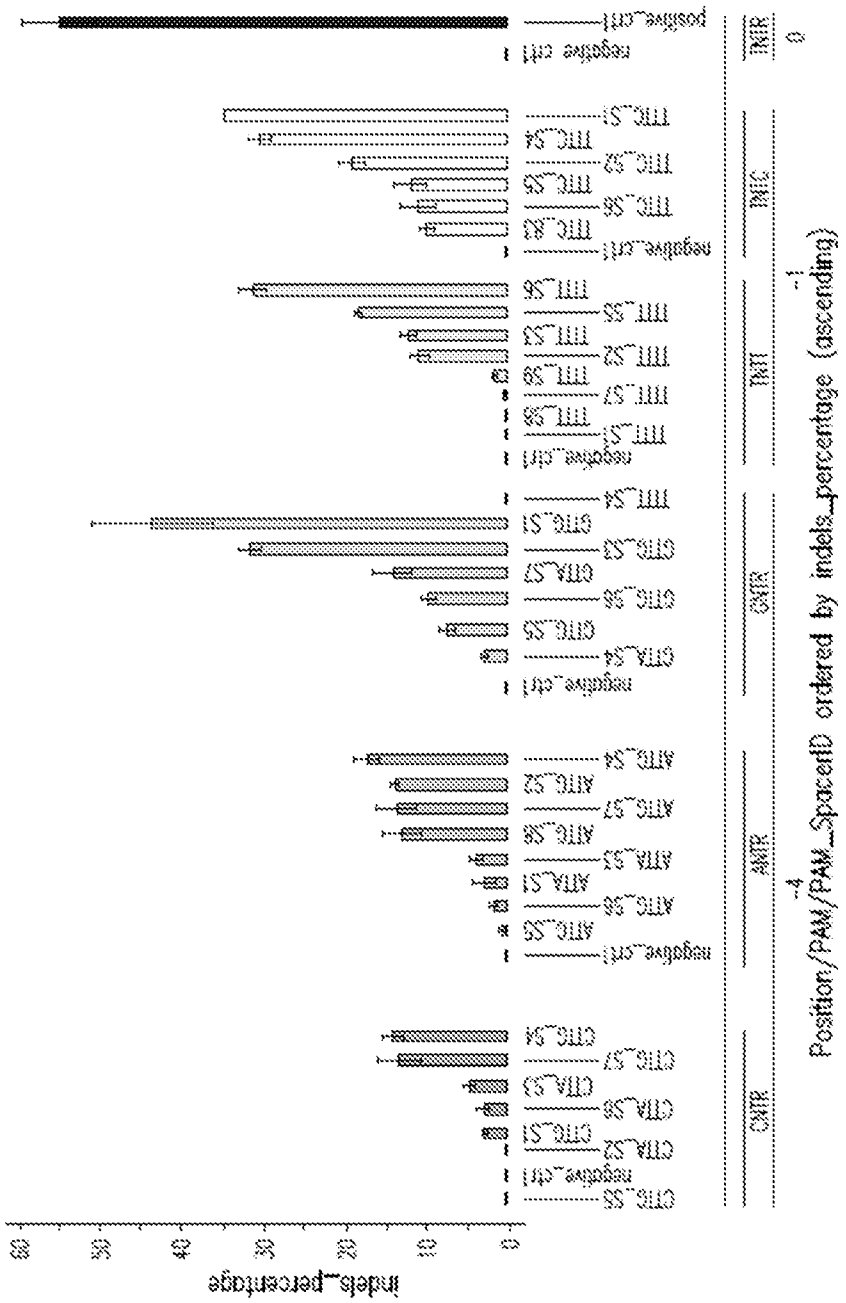
indels_percentage
Position/PAM/PAM_SpacerID ordered by indels_percentage (ascending)
-4
-1
0
5MTR
3MTR
5MTR
3MTR
5MTR
3MTR
positive_ctrl
negative_ctrl
TTTC_S1
TTTC_S4
TTTC_S2
TTTC_S5
TTTC_S6
TTTC_S3
negative_ctrl
TTTT_S6
TTTT_S5
TTTT_S3
TTTT_S2
TTTT_S9
TTTT_S7
TTTT_S8
negative_ctrl
TTTT_S1
TTTT_S4
GTTC_S1
GTTC_S2
GTTA_S7
GTTC_S6
GTTC_S5
GTTA_S4
negative_ctrl
ATTC_S4
ATTC_S2
ATTC_S7
ATTC_S8
ATTA_S3
ATTA_S1
ATTC_S6
ATTC_S5
negative_ctrl
CTTC_S4
CTTC_S7
CTTA_S5
CTTA_S6
CTTC_S1
CTTA_S2
negative_ctrl
CTTC_S3

FIG. 6B

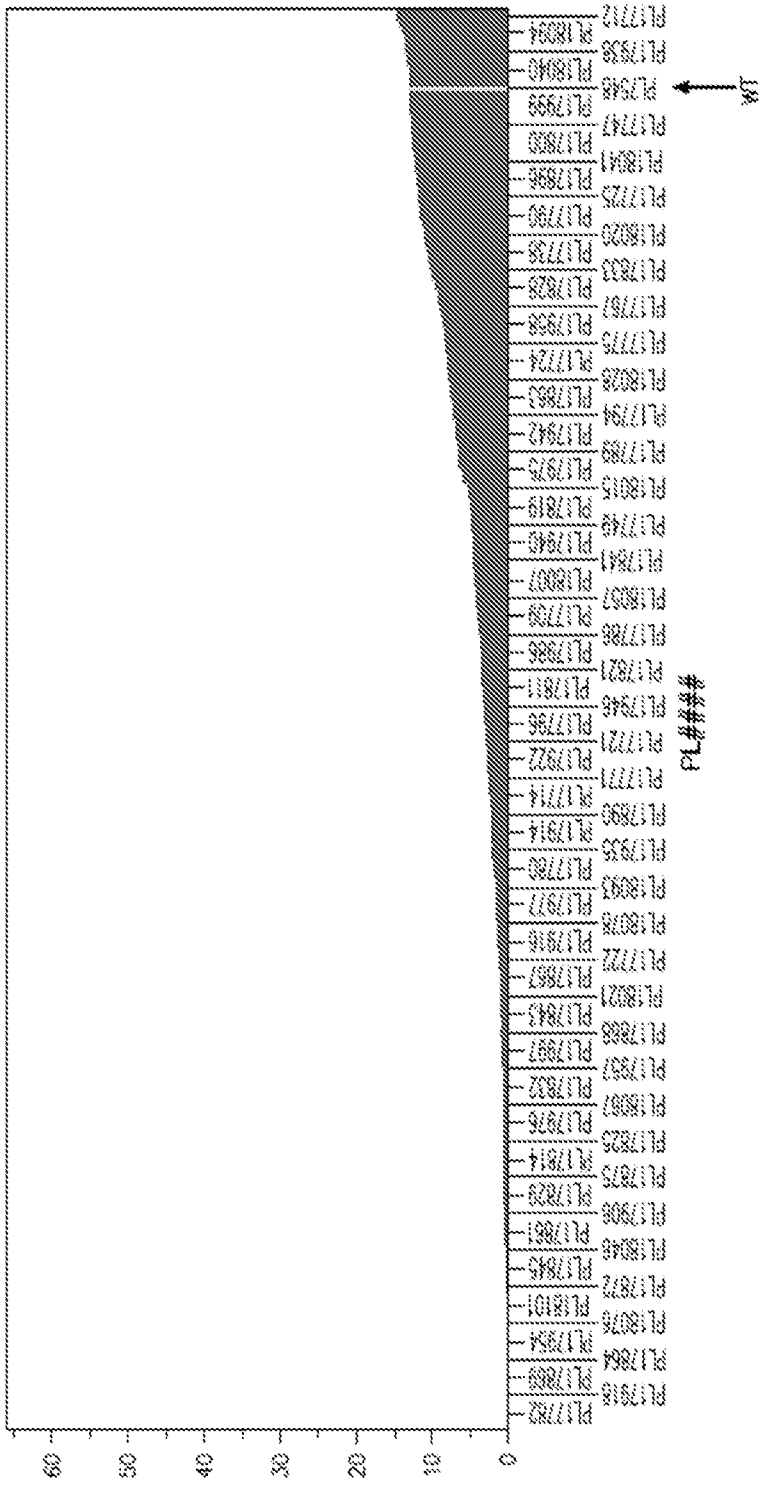

FIG. 7A

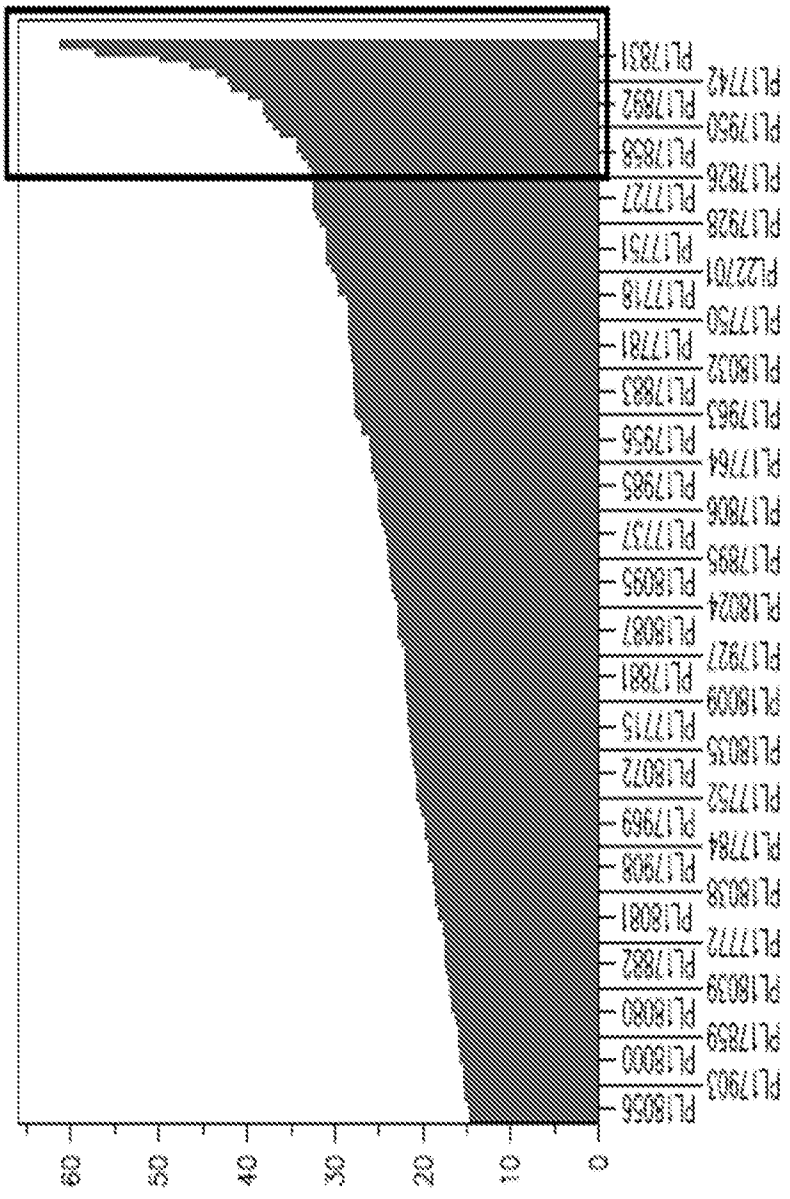
PL17831
PL17742
PL17892
PL17950
PL17838
PL17806
PL17727
PL17928
PL17751
PL22701
PL17718
PL17750
PL17781
PL18032
PL17823
PL17963
PL17936
PL17764
PL17965
PL17806
PL17737
PL17895
PL18095
PL18024
PL18087
PL17927
PL17861
PL18009
PL17715
PL18035
PL18072
PL17752
PL17969
PL17784
PL17908
PL18038
PL18081
PL17772
PL17862
PL18039
PL18080
PL17859
PL18000
PL17903
PL18056
60
50
40
30
20
10
0

FIG. 7B

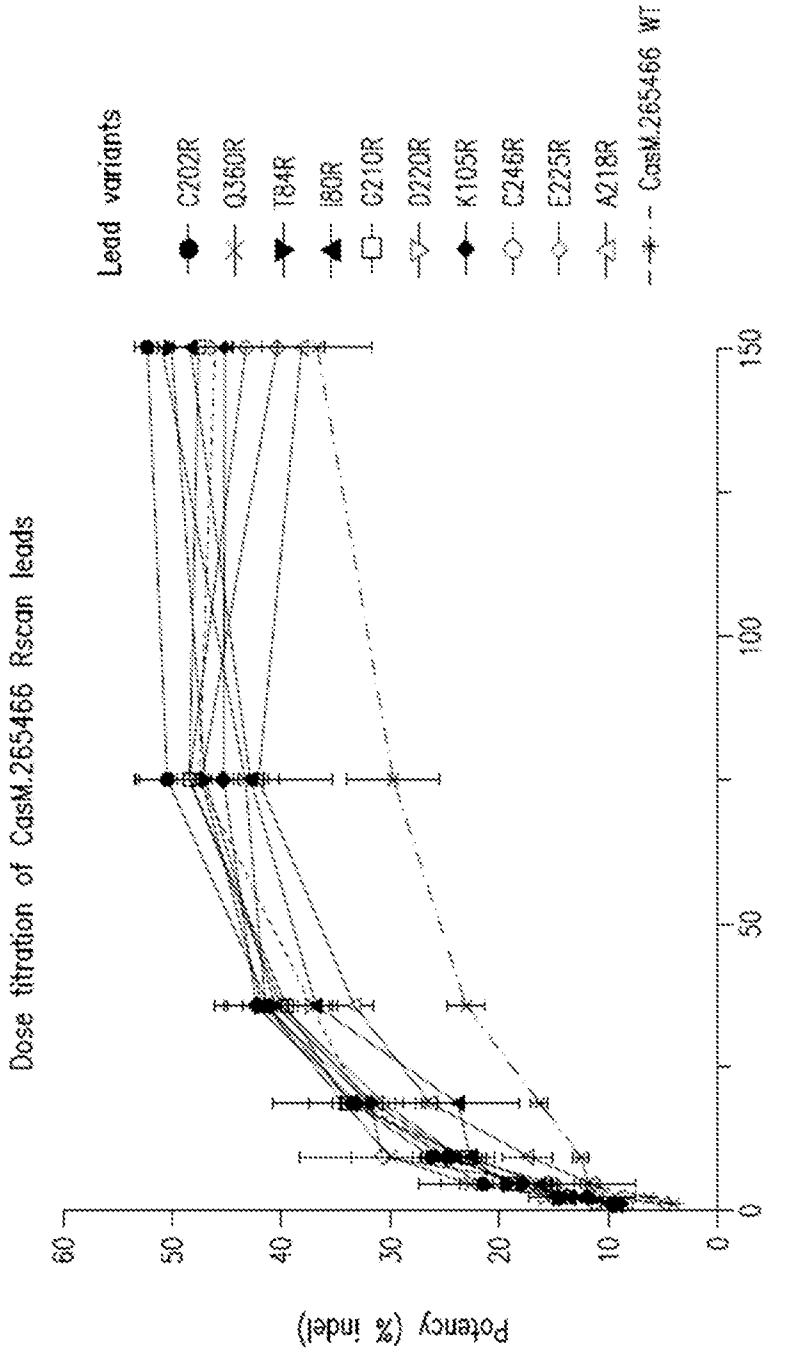
Dose titration of CasM.265466 Rscan leads
Lead variants
C202R
Q360R
T84R
I80R
C210R
D220R
K105R
C246R
E225R
A218R
CasM.265466 WT
Potency (% indel)
Dose (ng)
0
10
20
30
40
50
60
0
50
100
150

FIG. 8

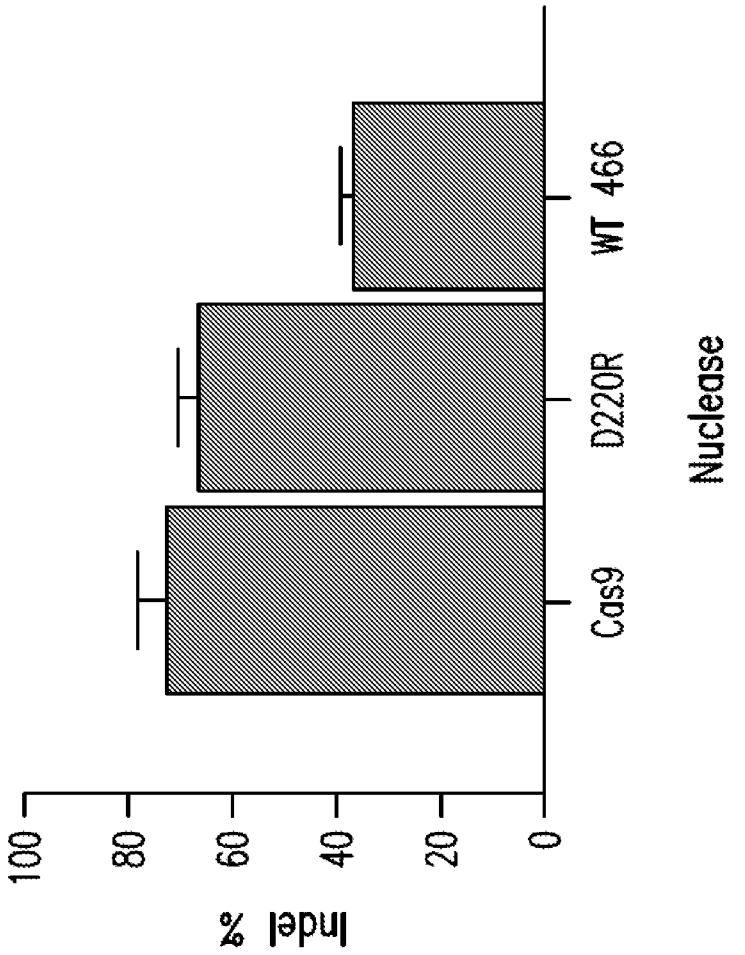

FIG. 9

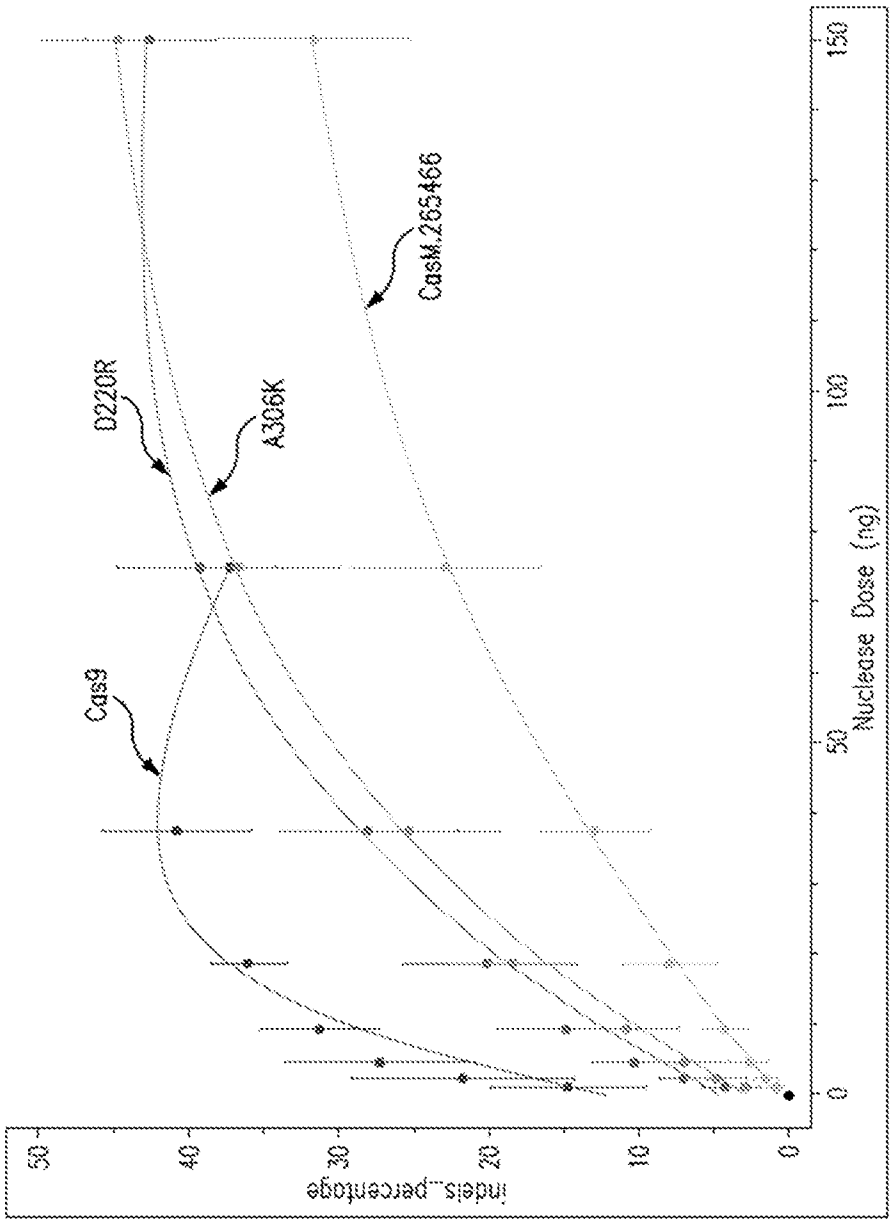
Cas9
D220R
A306K
CasM.265466
Nuclease Dose (ng)
indels_percentage
0
50
100
150
0
10
20
30
40
50

FIG. 10

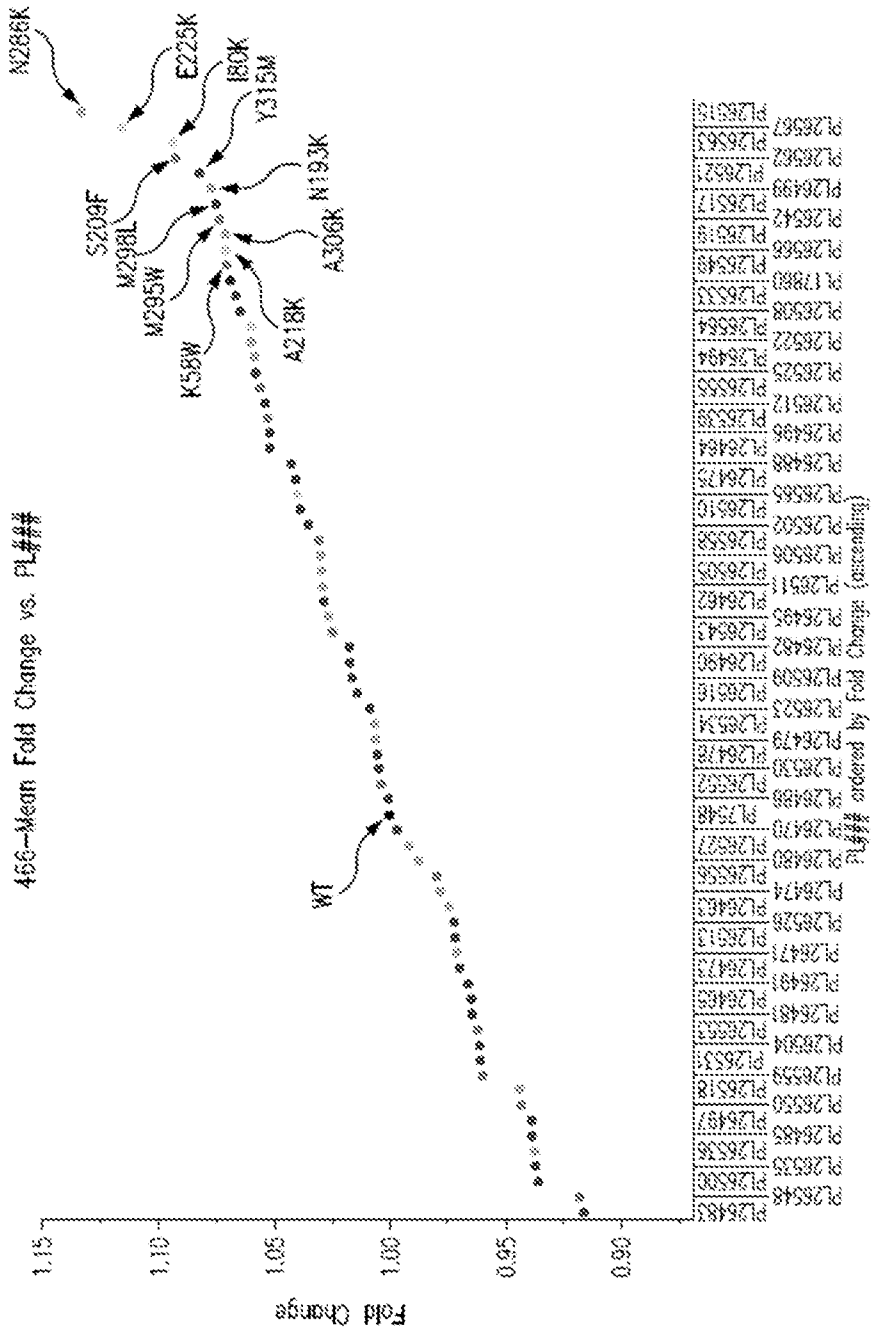
466-Mean Fold Change vs. PL###
Fold Change
1.15
1.10
1.05
1.00
0.95
0.90
WT
K58W
M295W
M298L
S209F
A218K
A306K
N193K
Y315M
I80K
E223K
M288K
PL### ordered by Fold Change (ascending)

FIG. 11

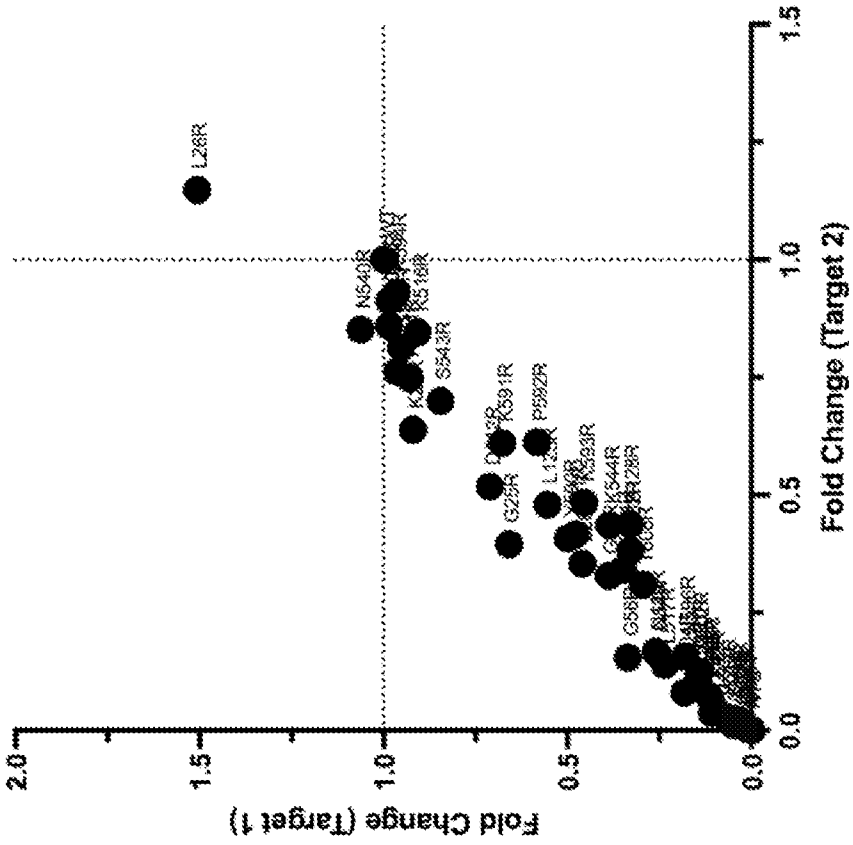
Fold Change (Target 1)
2.0
1.5
1.0
0.5
0.0
Fold Change (Target 2)
0.0
0.5
1.0
1.5
L26R
N54R
S54R
G2R
L12R

FIG. 12

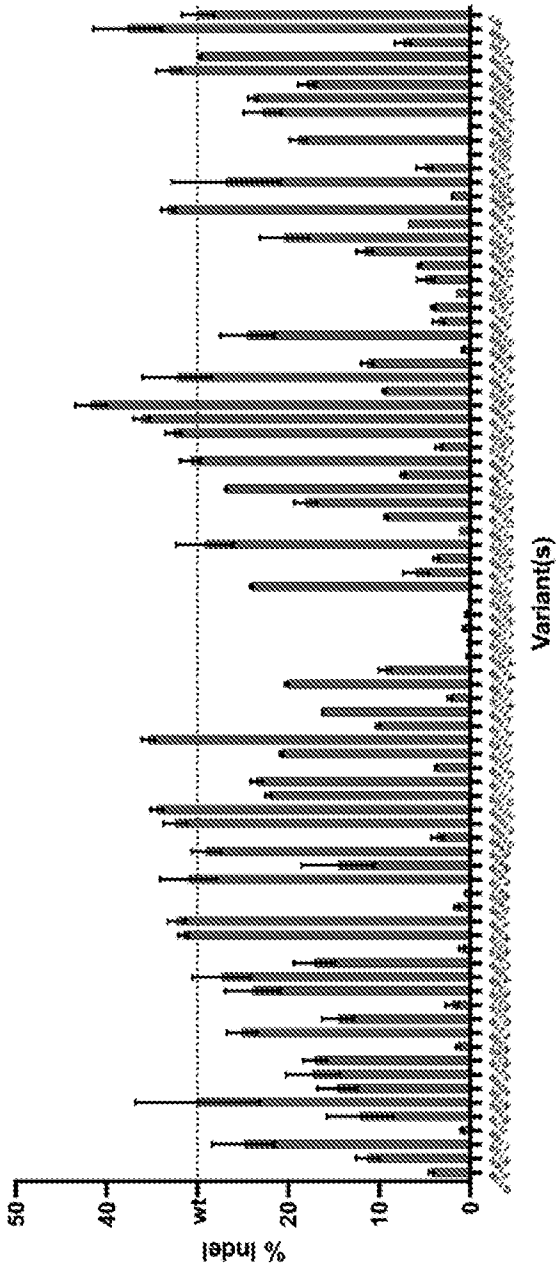

FIG. 13A

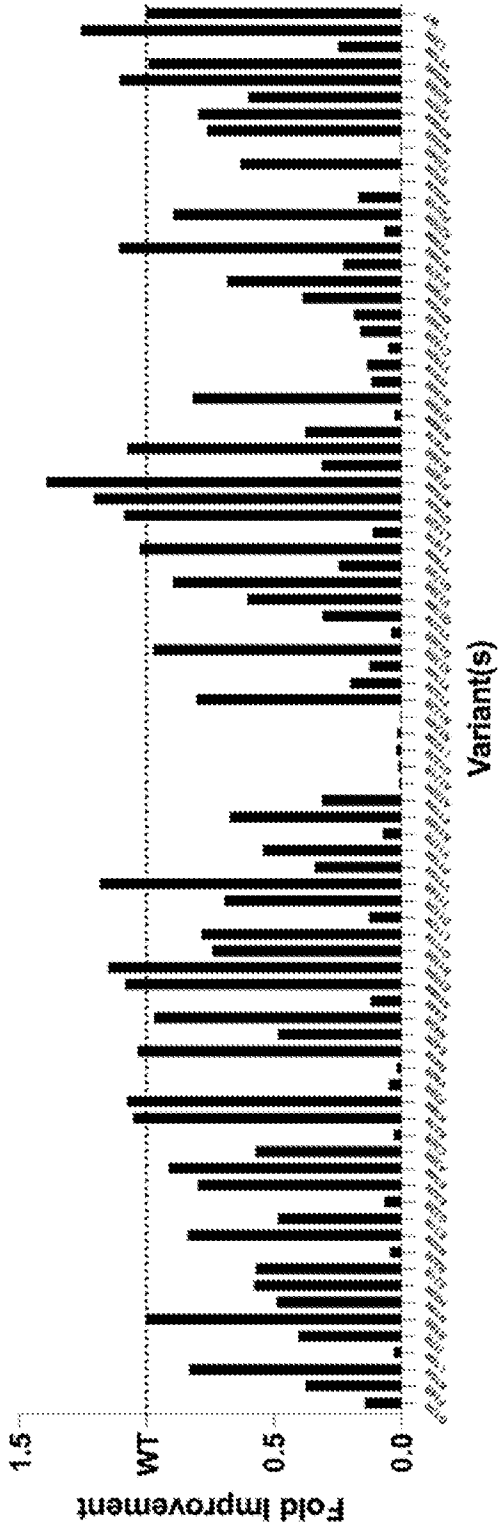

FIG. 13B

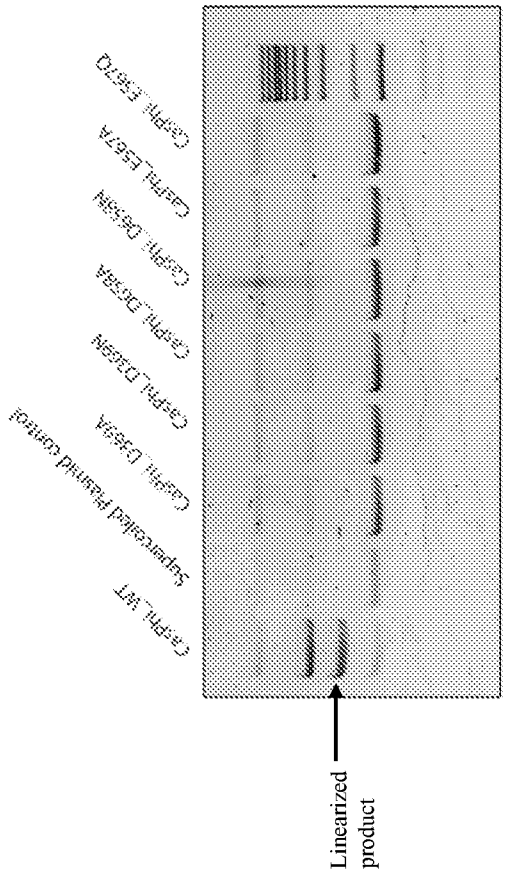
Linearized product
Cas9nL_WT
Supercoiled plasmid control
Cas9nL_D35A
Cas9nL_D35N
Cas9nL_D35R
Cas9nL_D35K
Cas9nL_E67A
Cas9nL_E67Q

FIG. 14

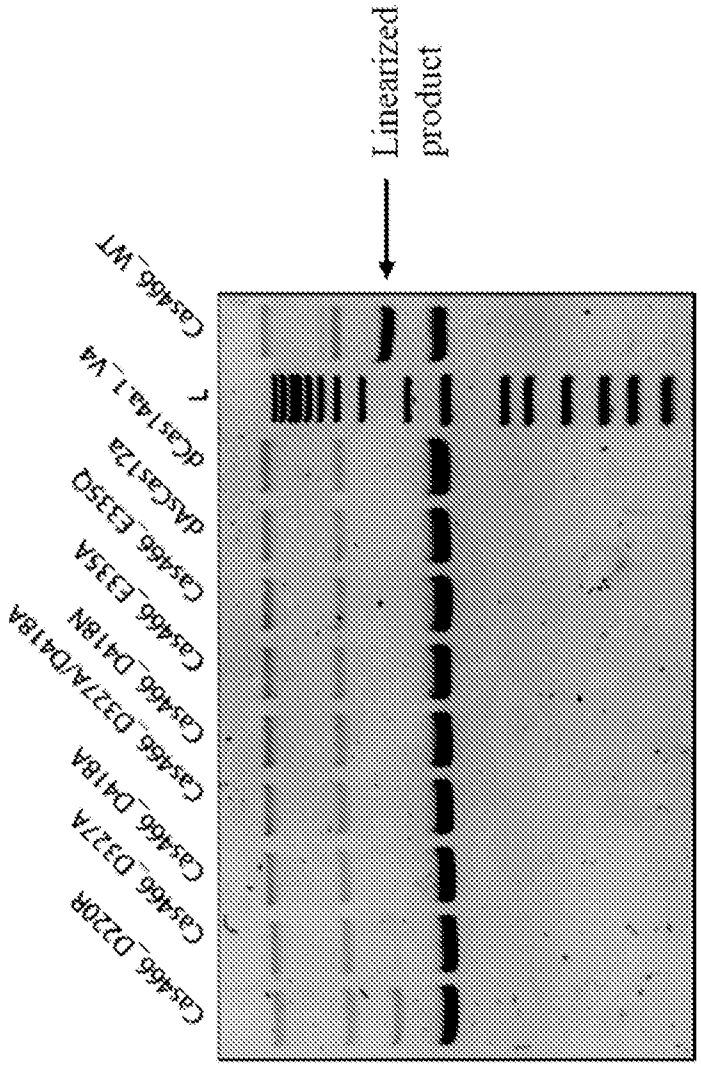
Linearized product
Cas466_WT
dCas5143.1_V4
dCas5123
dAsCas12a
Cas466_E335Q
Cas466_E335A
Cas466_D418N
Cas466_D327A/D418A
Cas466_D418A
Cas466_D327A
Cas466_D220R

FIG. 15

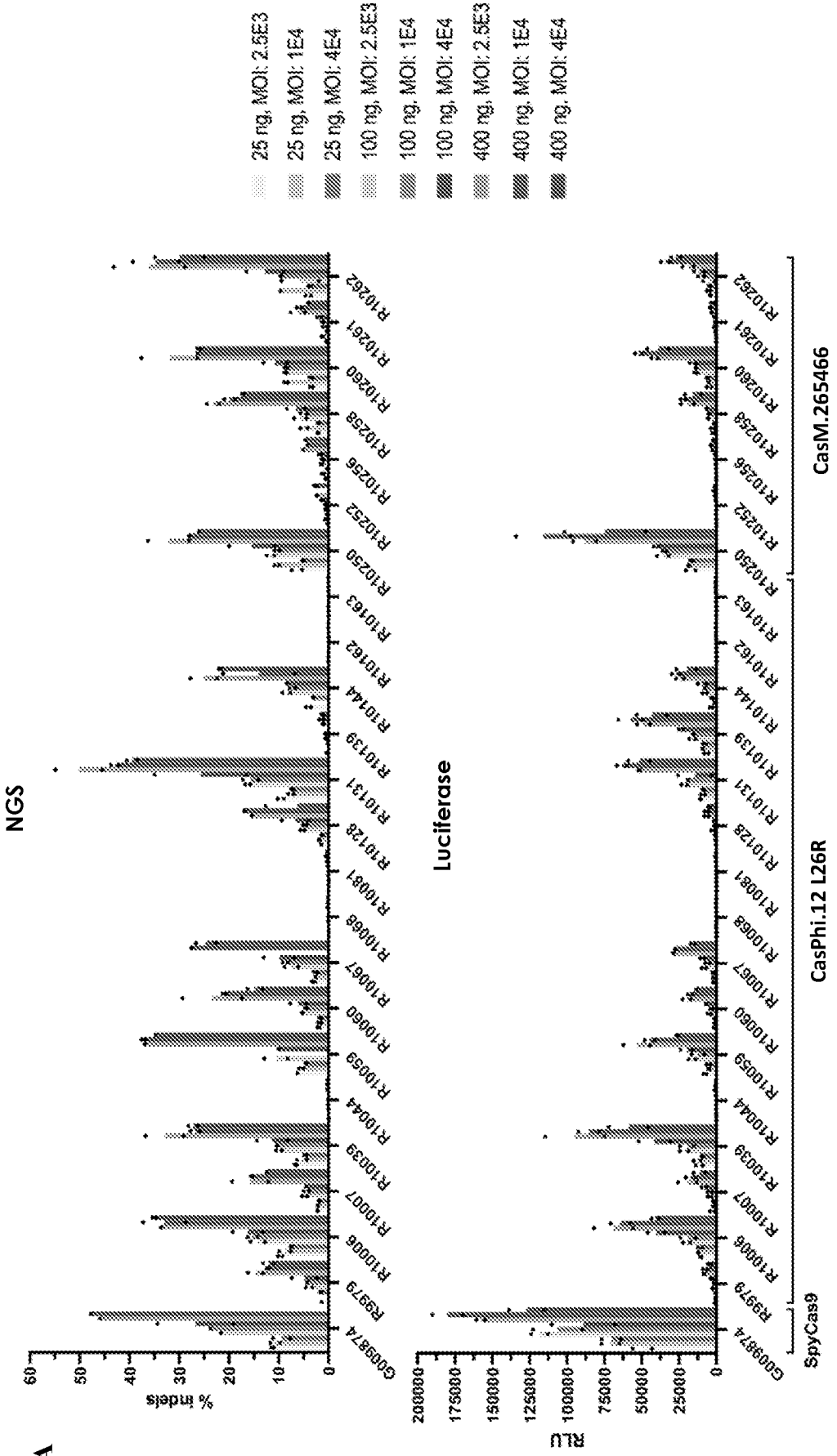

FIG. 16A

FIG. 16B

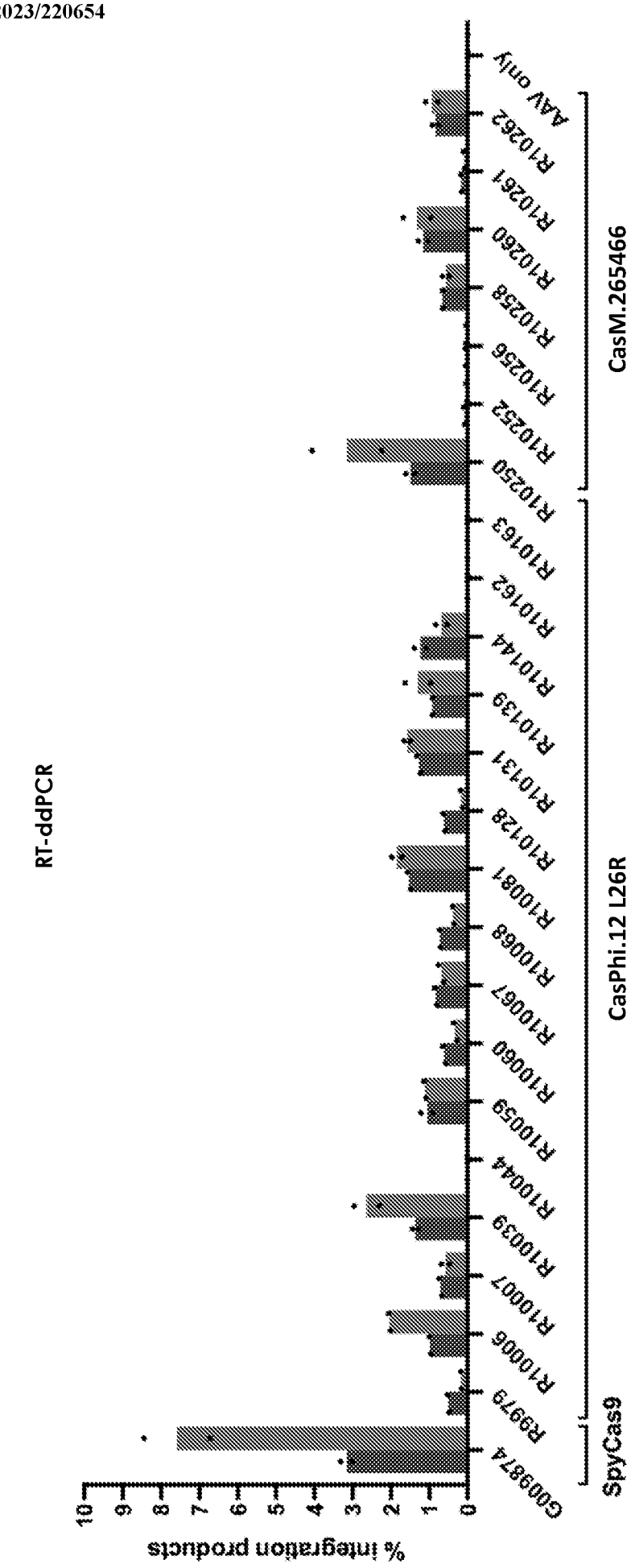
RT-ddPCR
% integration products
0
1
2
3
4
5
6
7
8
9
10
G009874
R9979
R10006
R10007
R10039
R10044
R10059
R10060
R10067
R10068
R10081
R10128
R10131
R10139
R10144
R10162
R10163
R10250
R10252
R10256
R10258
R10260
R10261
R10262
AAV only
SpyCas9
CasPhi.12 L26R
CasM.265466

FIG. 17